# 2 Các hệ thống máy tính

## Tổng quan

Một hệ thống máy tính gồm có phần cứng và phần mềm. Có nhiều loại máy tính, nhưng nguyên tắc hoạt động của chúng về cơ bản là giống nhau. Chúng ta sẽ học về cấu tạo của máy tính (phần cứng) trong phần 1 và phần mềm (hệ điều hành) cho các máy tính hiệu năng cao trong phần 2. Chúng ta cũng sẽ học về cấu hình của hệ thống máy tính cho việc cải thiện tính tin cậy trong phần 3 và cách đánh giá hiệu năng của máy tính trong phần 4. Cuối cùng trong phần 5 chúng ta sẽ học về các hệ thống mà sử dụng máy tính.

2.1 Phần cứng
2.2 Hệ điều hành
2.3 Công nghệ cấu hình hệ thống
2.4 Hiệu năng và độ tin cậy của hệ thống
2.5 Các hệ thống ứng dụng

**[Các thuật ngữ và khái niệm cần biết]**
Đơn vị xử lý trung tâm (CPU), bộ nhớ đệm, giao tiếp vào/ra, bộ nhớ hỗ trợ, quản lý tác vụ, quản lý công việc, đa lập trình, bộ nhớ ảo, hệ thống kép, hệ thống song công, hệ thống khách/chủ, tính sẵn sàng, MTBF, MTTR, Internet

# 2.1 Phần cứng

## Giới thiệu

Phần cứng là các thiết bị cấu tạo nên máy tính. Máy tính bao gồm khối xử lý, bộ nhớ, các khối vào ra... Phần này sẽ xem xét các khối này trên quan điểm của phần cứng.

## 2.1.1 Các phần tử thông tin (Bộ nhớ)

| Điểm chính | ➢ Phần tử thông tin gồm có ROM và RAM<br>➢ SRAM và DRAM là hai loại tiêu biểu của RAM |
|---|---|

**Bộ nhớ bán dẫn** là bộ nhớ được làm từ các mạch tích hợp (ICs) sử dụng chất liệu bán dẫn. Bộ nhớ bán dẫn bao gồm **ROM** bộ nhớ không thể ghi lại được và **RAM** bộ nhớ có thể ghi lại được

### ROM (Bộ nhớ chỉ đọc)

**ROM** là bộ nhớ bán dẫn mà dữ liệu không bị xóa khi mất điện[1]. Với Mask ROM và PROM thì dữ liệu chỉ được ghi một lần; tuy nhiên EPROM dữ liệu có thể được ghi nhiều lần sử dụng một phương pháp đặc biệt. Phân loại và đặc điểm của ROM được trình bày trong bảng dưới đây

| Loại (Tên) | Đặc điểm |
|---|---|
| Mask ROM | Dữ liệu được ghi vào tại thời điểm sản xuất. Dữ liệu không thể ghi lại được sau đó. |
| PROM (ROM lập trình được) | Dữ liệu PROM được ghi bởi người dùng trong lần đầu tiên sử dụng. Dữ liệu không thể ghi lại được ở các lần sau. |
| EPROM (PROM có thể xóa) | Dữ liệu EPROM được ghi bởi người dùng bằng điện. Toàn bộ dữ liệu có thể xóa được bằng cách sử dụng tia tử ngoại. |
| EEPROM (Electrically EPROM) | Toàn bộ dữ liệu có thể xóa và ghi lại được. Dữ liệu được xóa bằng điện. |
| Flash memory[2] (Bộ nhớ cực nhanh) | Việc xóa và ghi lại được thực hiện một cách tập trung hoặc theo khối. Dữ liệu được xóa bằng điện. |

[1] Tính khả biến và tính bất biến: Thuộc tính làm cho nội dung của bộ nhớ bị mất khi mất điện được gọi là tính khả biến. RAM là loại bộ nhớ có tính khả biến. Ngược lại thuộc tính làm cho nội dung của bộ nhớ không bị mất khi mất điện gọi là tính bất biến. ROM là loại bộ nhớ có tính bất biến.

[2] (Gợi ý) Bộ nhớ Flash được phân loại như EEPROM.

# RAM (Bộ nhớ truy cập ngẫu nhiên)

**RAM** là bộ nhớ bán dẫn mà nội dung của nó sẽ bị mất khi mất điện. Khác với ROM, nội dung của RAM có thể thay đổi được vì vậy nó được dùng cho bộ nhớ chính, bộ nhớ đồ họa[3] và bộ nhớ đệm.

Hai loại RAM thông thường là **SRAM** và **DRAM.** Đặc điểm của SRAM và DRAM được tổng hợp và đưa ra trong bảng dưới đây

| Các mục so sánh | SRAM | DRAM |
|---|---|---|
| Mức độ tích hợp | Thấp (dung lượng nhỏ) | Cao (dung lượng lớn) |
| Tốc độ truy cập | Nhanh | Chậm |
| Giá thành | Đắt | Không đắt |
| Cách dùng | Bộ nhớ đệm[4], Các thiết bị hoạt động bằng pin | Bộ nhớ chính |
| Hoạt động | Không cần phải làm tươi | Cần phải làm tươi. |
| Cấu trúc | Flip-flop<br>Cấu trúc phức tạp | Tụ điện và các transistors<br>Cấu trúc đơn giản |

### SRAM (RAM tĩnh)

SRAM được tạo thành từ các flip-flop[5], vì vậy không cần thao tác làm tươi và có thể tăng tốc độ đọc và ghi thông tin. Tuy nhiên giá thành SRAM cao hơn DRAM với cùng một dung lượng do cấu trúc của SRAM phức tạp hơn so với DRAM. Vì lý do này nó được dùng chủ yếu ở những nơi cần tốc độ cao, không quan trọng giá thành, những nơi quan trọng, ví dụ như trong bộ nhớ đệm. SRAM cũng thường được sử dụng trong các thiết bị hoạt động bằng pin.

### DRAM (RAM động)

DRAM được tạo thành từ các tụ điện và các transistor, biểu diễn việc có hoặc không có điện tích trong tụ điện bởi các logic 1 và 0. Sau một thời gian thì điện tích trong tụ điện sẽ bị mất đi, đó là nguyên nhân dẫn đến mất dữ liệu. Do đó DRAM cần được ghi lại (làm tươi lại) trong những khoảng thời gian nhất định (cứ mỗi vài mili giây). Vì cấu trúc của DRAM đơn giản hơn, giá thành sản xuất thấp nên được sử dụng chủ yếu trong bộ nhớ chính của máy tính PC[6].

Các loại DRAM được trang bị chức năng trao đổi dữ liệu tốc độ cao gồm có SDRAM và DDRAM

[3] **Bộ nhớ đồ họa**: Đó là bộ nhớ được sử dụng khi ảnh hay ký tự được hiển thị trên màn hình sử dụng một máy tính. Nó cũng được gọi là bộ nhớ video (VRAM).
[4] **Bộ nhớ đệm**: Đó là bộ nhớ tốc độ cao nằm giữa bộ nhớ chính và CPU để tăng tốc độ đọc dữ liệu từ bộ nhớ chính đến CPU.
[5] **Flip-flop** (còn gọi là mạch 2 trạng thái): Là một mạch điện với 2 trạng thái ổn định, nó giữ các trạng thái đó cho đến khi có một đầu vào làm thay đổi một trong các trạng thái đó.
[6] (FAQ) Thường xuyên có những câu hỏi so sánh giữa SRAM và DRAM. Bạn cần biết về mức độ tích hợp, cách sử dụng, cấu trúc...

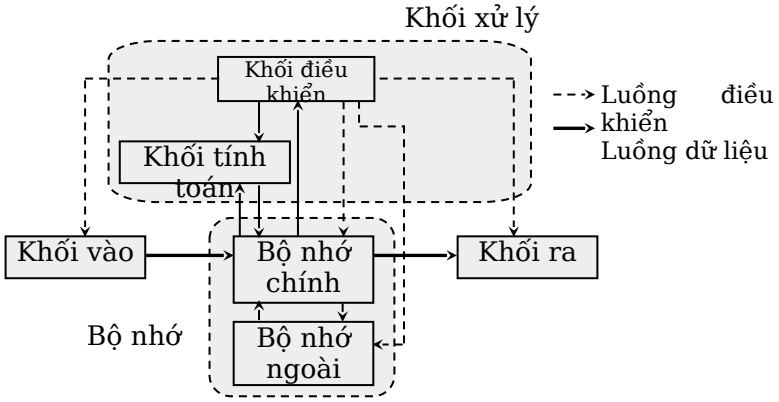
# 2.1.2 Kiến trúc bộ xử lý

| Điểm chính | ➢ Máy tính bao gồm 5 khối chính<br>➢ Có rất nhiều phương pháp định địa chỉ như định địa chỉ trực tiếp, định địa chỉ gián tiếp... |
|---|---|

Thuật ngữ kiến trúc ở đây còn liên quan đến "cấu trúc hay tổ chức". Kiến trúc bộ xử lý cũng liên quan đến cấu hình và các hoạt động cơ bản của máy tính

## Cấu hình của máy tính

Hình dưới đây chỉ ra cấu hình cơ bản của một máy tính. Cấu hình này được gọi là "5 khối lớn" hay "5 chức năng lớn"[7] bởi vì có 5 thành phần chính




Khối điều khiển và khối tính toán kết hợp với nhau gọi là khối xử lý hay khối xử lý trung tâm (CPU).

## Sự tính toán địa chỉ và các phương pháp định địa chỉ

Một **chương trình** được lưu trong bộ nhớ chính và được thực hiện theo từng lệnh bởi khối điều khiển. Sự tính toán địa chỉ để xác định vị trí của dữ liệu cần được xử lý.

---

[7] **"5 khối lớn"**:

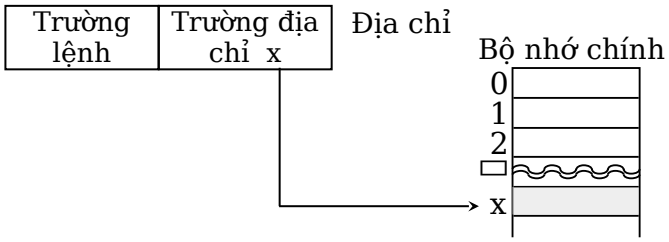
- **Khối điều khiển**: Là khối điều khiển toàn bộ máy tính. Nó đọc lệnh từ bộ nhớ chính và gửi đến các khối cần thiết để thực hiện lệnh
- **Khối tính toán**: Là khối thực hiện các phép tính toán số học, các phép tính toán logic, và các phép tính toán khác. Nó bao gồm các bộ cộng, các thanh ghi, các bộ đảo (đơn vị mà chuyển đổi các giá trị sang giá trị đảo của chúng), ...
- **Bộ nhớ**: Thuật ngữ chung của các khối lưu trữ dữ liệu, chương trình, ... Được phân loại thành bộ nhớ chính và bộ nhớ ngoài (bộ nhớ phụ).
- **Khối vào**: Thuật ngữ chung của các khối mà nhập chương trình và dữ liệu vào máy tính.
- **Khối ra**: Thuật ngữ chung của khối mà đưa ra các kết quả xử lý của máy tính dưới dạng kí tự hoặc con số mà chúng ta có thể nhận biết được.

**Sự tính toán địa chỉ** là chức năng lấy về địa chỉ thực dựa trên trường địa chỉ cho trong lệnh. Phương pháp tính toán địa chỉ được gọi là phương pháp **định địa chỉ.** Địa chỉ truy cập thực sự, kết quả của phép tính toán địa chỉ được gọi là **địa chỉ hiệu dụng**. Các phương pháp định địa chỉ được mô tả chi tiết dưới đây

### Phương pháp định địa chỉ trực tiếp

Trong phương pháp này, nội dung ô nhớ được lưu trữ trong trường địa chỉ của lệnh là dữ liệu cần thao tác.

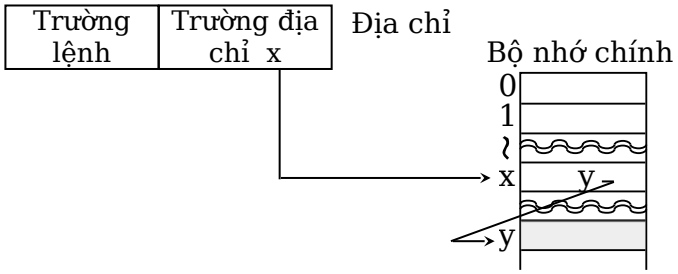
Phương pháp định địa chỉ trực tiếp




### Phương pháp định địa chỉ gián tiếp

Trong phương pháp này, nội dung ô nhớ được lưu trữ trong trường địa chỉ của lệnh chưa phải là dữ liệu cần thao tác; Mà dữ liệu được lưu trữ tại địa chỉ được chỉ ra bởi nội dung đó mới là dữ liệu cần thao tác.

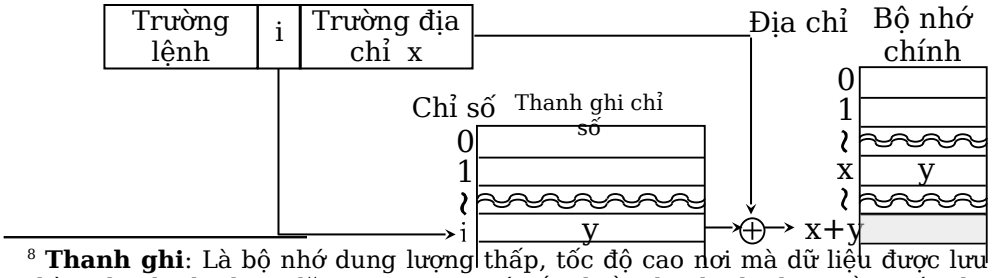
Phương pháp định địa chỉ gián tiếp




### Phương pháp định địa chỉ chỉ số (thay đổi chỉ số)

Trong phương pháp này, địa chỉ hiệu dụng được tính bằng tổng giá trị của địa chỉ (lưu trong trường địa chỉ của lệnh) với giá trị của thanh ghi chỉ số. [8] Ví dụ, khi xử lý một mảng, chúng ta có thể tham chiếu nội dung của các địa chỉ khác đơn giản bằng cách thay đổi nội dung của thanh ghi chỉ số. [9]

Phương pháp định địa chỉ chỉ số

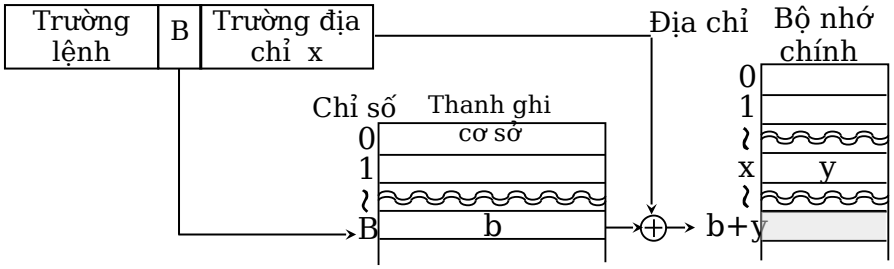



[8] **Thanh ghi**: Là bộ nhớ dung lượng thấp, tốc độ cao nơi mà dữ liệu được lưu trữ một cách tạm thời. Thanh ghi được đặt trong CPU. Có rất nhiều thanh ghi, bao gồm: các thanh ghi chung, cho việc lưu trữ tức thời và kết quả cuối cùng của các phép toán; thanh ghi trạng thái chỉ ra trạng thái của CPU sau khi một lệnh được thực hiện; thanh ghi chỉ số cho việc tính toán địa chỉ; và thanh ghi cơ sở.

[9] (FAQ) Có nhiều câu hỏi về khái niệm của các phương pháp định địa chỉ. Ví dụ "Câu nào trong các câu sau là mô tả phù hợp của phương pháp định địa chỉ trực tiếp?" hãy chắc chắn rằng bạn đã nắm vững các phương pháp này: phương pháp định địa chỉ trực tiếp, phương pháp định địa chỉ chỉ số, phương pháp định địa chỉ tức thì...

### Phương pháp định địa chỉ cơ sở

Trong phương pháp này, địa chỉ hiệu dụng là tổng của địa chỉ được chỉ ra bởi trường địa chỉ của lệnh và nội dung của thanh ghi địa chỉ cơ sở[10]
Phương pháp định địa chỉ cơ sở




### Phương pháp định địa chỉ tức thì

Theo phương pháp này thì trường địa chỉ của lệnh lưu trữ trực tiếp dữ liệu cần xử lý, không phải là địa chỉ. [11]

# RISC và CISC

Có rất nhiều loại công nghệ đã được sử dụng để tăng tốc độ xử lý máy tính; RISC và CISC là những ví dụ của công nghệ này. Với sự tiến bộ của công nghệ bán dẫn, mật độ tích hợp của các mạch tích hợp tăng lên một cách liên tục. Hiện tại, máy tính được làm từ các mạch tích hợp, và RICS và CISC là 2 cách tiếp cận cho việc phát triển các máy tính. Các máy tính được cấu hình cho tốc độ cao với tập lệnh đơn giản và phần cứng đã được đơn giản hóa được gọi là RISCs. Ngược lại cấu trúc tập lệnh phức tạp được cấu hình trong một mạch được gọi là CISCs.

### RISC (Máy tính với tập lệnh rút gọn)

Các máy tính này chỉ có một tập lệnh đơn giản, các lệnh thường xuyên sử dụng được tích hợp trong một chip VLSI (tích hợp mật độ rất lớn) để đạt được hiệu năng cao thông qua việc cải thiện tốc độ xử lý và giảm thời gian thực hiện lệnh. Điểm đáng lưu ý của kiến trúc RISC là độ dài của của các lệnh là cố định và như nhau, thời gian thực hiện mỗi lệnh cũng được hạn chế. Bằng cách làm như vậy dễ dàng áp dụng điều khiển đường ống lệnh(pipeline). Tuy nhiên, số lượng các lệnh phải thực hiện trở nên lớn trừ khi các chương trình đối tượng có hiệu suất cao được tạo, vì vậy trình biên dịch (compiler) cần thiết phải có chức năng tối ưu hóa[12]. Hầu hết các máy

[10] (Chú ý) Phương pháp định địa chỉ cơ sở có thể được sử dụng không cần chú ý tới chương trình được lưu ở đâu trong bộ nhớ, đơn giản bằng cách thay đổi giá trị của thanh ghi cơ sở. Một cấu trúc như thế được gọi là cấu trúc có thể xác định lại vị trí.
[11] (Gợi ý) Chú ý rằng trong phương pháp định địa chỉ tức thì không chỉ định địa chỉ của bộ nhớ chính.
[12] **Sự tối ưu hóa**: Đó là một chức năng của trình biên dịch bỏ đi những phần thừa của một chương trình để giảm thời gian thực thi của chương trình và kích cỡ của chương trình. Điều này được thực hiện bằng nhiều cách, ví dụ như tính toán hằng số, đơn giản hóa các công thức, và giảm các vòng lặp lồng nhau.

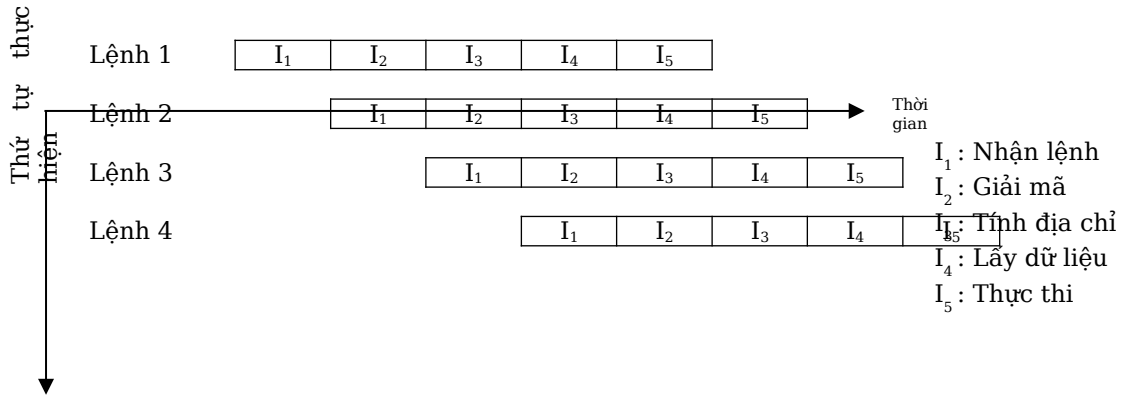
tính trạm đều thuộc loại này.

### CISC (Máy tính với tập lệnh phức tạp)

Những máy tính này có tập lệnh phức tạp được tích hợp trên một VLSI chip theo để đạt được hiệu năng toàn bộ cao. Hầu hết các máy tính mục đích chung đều có kiến trúc CISCs.

## Điều khiển đường ống lệnh

Chúng ta đã đề cập RISC và CISC như là những công nghệ để cải thiện tốc độ xử lý của máy tính. Để cải thiện hơn nữa tốc độ xử lý, hệ thống RISC sử dụng điều khiển đường ống lệnh. Điều khiển đường ống lệnh là một kỹ thuật giảm thời gian thực hiện lệnh của CPU. Việc này được thực hiện như sau: Các bước thực hiện của một lệnh được chia ra làm 5 hoặc 6 bước nhỏ, và nếu mỗi bước được hoàn thành trong một khoảng thời gian xác định và các bước lệnh là độc lập với các bước khác thì chúng ta có thể cải thiện tốc độ xử lý bằng cách trễ việc thực hiện mỗi lệnh một bước so với lệnh trước đó. Tuy nhiên trong thực tế, việc ghép nối lệnh, sẽ mất thời gian khi địa chỉ lệnh tiếp theo chưa được xác định, và một vài bước không thể hoàn thành trong một khoảng thời gian cố định. Điều khiển đường ống lệnh xử lý lệnh một cách đồng thời và cung cấp một phương pháp để tăng tốc độ của máy tính mặc dù mọi việc không phải lúc nào cũng hoạt động theo đúng ý tưởng [13]




# 2.1.3 Kiến trúc bộ nhớ

| Điểm chính | ➢ Cấu trúc bộ nhớ phân cấp được đưa ra để đạt được tốc độ cao và dung lượng lớn<br>➢ Bộ nhớ đệm và kỹ thuật đan xen được sử dụng để tăng tốc độ truy cập bộ nhớ |
|---|---|

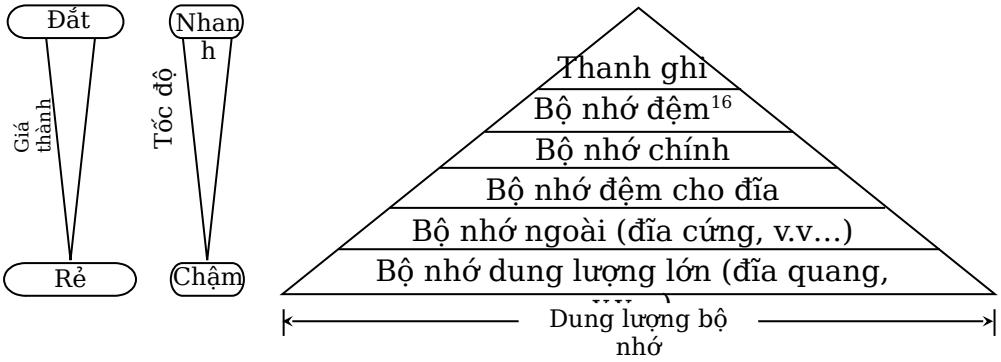
Có rất nhiều yêu cầu đối với bộ nhớ, nhưng yêu cầu về tốc độ cao và dung lượng lớn[14] là những yêu cầu quan trọng. Tuy nhiên, nói chung là bộ nhớ tốc độ cao thì thường đắt và có dung lượng nhỏ trong khi bộ nhớ tốc độ thấp thì giá thành thấp và có dung lượng cao hơn. Như vậy việc cần làm là kết hợp bộ nhớ tốc độ cao nhưng dung lượng nhỏ với bộ nhớ tốc độ thấp nhưng dung lượng lớn để tạo ra bộ nhớ có tốc độ cao và dung lượng lớn.

[13] (FAQ) Các câu hỏi về điều khiển đường ống lệnh thường xuất hiện trong các bài thi. Hầu hết chúng đều ở dạng chọn một mô tả phù hợp với điều khiển đường ống lệnh, vì vậy bạn chỉ cần biết điều khiển đường ống lệnh thực thi lệnh một cách đồng thời.

[14] (Chú ý) Bên cạnh đó, yêu cầu đối với bộ nhớ bao gồm tính tin cậy, khả năng truy cập ngẫu nhiên, tính bất biến, chức năng có thể ghi lại được, khả chuyển, giá thành thấp…

# Phân cấp bộ nhớ

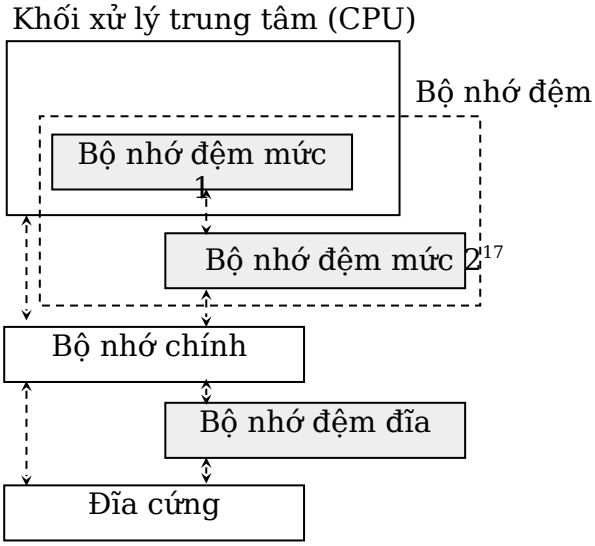
Phân cấp bộ nhớ là biểu diễn phân cấp của mối quan hệ giữa tốc độ truy cập và dung lượng lưu trữ của các loại bộ nhớ.[15]




# Bộ nhớ đệm nhanh (cache)

**Bộ nhớ đệm (cache)** là bộ nhớ tốc độ cao, dung lượng thấp, nằm giữa CPU hoặc thanh ghi và bộ nhớ chính. Tốc độ của bộ nhớ chính thấp hơn so với tốc độ của CPU và thanh ghi, do đó để tăng hiệu quả xử lý của CPU thì các dữ liệu và chương trình truy cập thường xuyên trong bộ nhớ chính được nạp vào bộ nhớ đệm. Gần đây, người ta đưa thêm một bộ nhớ đệm phụ vào để cải thiện hơn nữa tốc độ truy cập bộ nhớ.

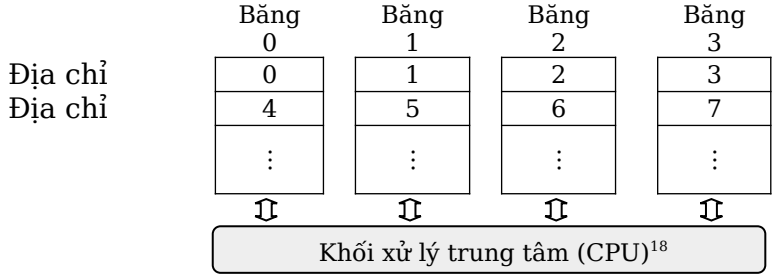
Bộ nhớ đệm nằm giữa ổ cứng và bộ nhớ chính được gọi là **bộ đệm đĩa.**
Hình dưới đây mô tả mối quan hệ giữa bộ nhớ đệm và bộ đệm đĩa.




[15] (Chú ý) Nếu $t$ là thời gian truy cập bộ nhớ trung bình, $t_m$ là thời gian truy cập bộ nhớ chính, $t_c$ là thời gian truy cập bộ nhớ đệm, và $h$ là tỷ lệ hit, thì ta có phương trình sau: $t = t_c h + t_m(1 - h)$.
[16] **Tỉ lệ hit**: Đó là xác suất mà phần chương trình cần thực thi chương trình nằm trong bộ nhớ đệm.
[17] **Bộ nhớ đệm mức 2**: Bộ nhớ đệm chính là bộ nhớ đệm được xây dựng trong CPU; Bộ nhớ đệm thứ 2 là bộ nhớ đệm nằm giữa đệm chính và bộ nhớ chính.

## Kỹ thuật đan xen (Tổ chức bộ nhớ đan xen)

Bộ nhớ chính được chia thành rất nhiều khối nhỏ gọi là các băng, và địa chỉ được gán qua các băng. Thông thường bộ nhớ chính được truy cập thông qua một dải tuần tự của địa chỉ ở một thời điểm do đó tốc độ có thể được tăng cường bằng cách truy cập đồng thời một dải tuần tự địa chỉ. Ví dụ ngay cả khi băng 0 chưa có dữ liệu thì băng 1 vẫn có thể truy cập. Hình dưới đây là một ví dụ về đan xen 4 (với 4 băng nhớ)




Dữ liệu và chương trình được lưu trữ trên một dãy địa chỉ (theo chiều ngang), nhưng bộ nhớ được truy cập theo các băng nhớ (theo chiều dọc). Điều này cho phép đồng thời truy cập đến một dải địa chỉ.

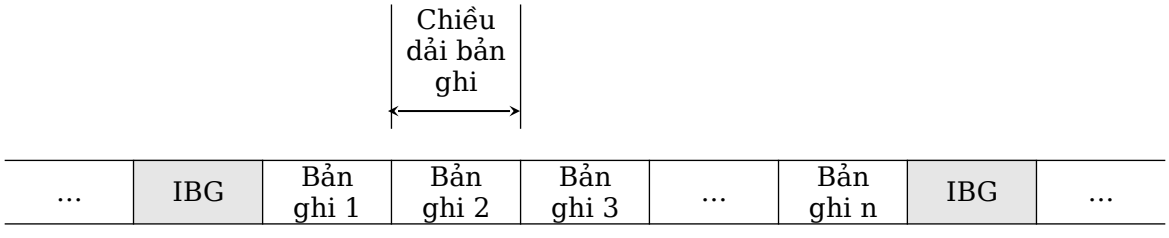
# 2.1.4 Các đơn vị băng từ

| **Điểm chính** | ➢ Sự tính toán dung lượng theo hệ số khối (block) và mật độ bản ghi<br>➢ Sự tính toán hiệu năng không bao gồm thời gian dừng |
|---|---|

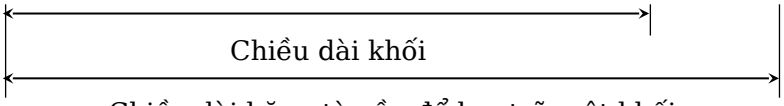
**Băng từ** là phương tiện lưu trữ dữ liệu trên một băng đã được nhiễm từ tính. Giá của phương tiện nhớ này rẻ và có dung lượng lớn, vì vậy nó được dùng cho các thiết bị như ổ đĩa cứng.

## Tính toán dung lượng

Định dạng bản ghi của băng từ được chỉ ra dưới đây. Ta có thể nhìn thấy trong hình này, để lưu một khối cần thêm vào một IBG (khoảng cách giữa các khối) là một vùng để xác định khối đó và chứa mã đặc biệt. Đầu đọc băng từ đọc dữ liệu theo các khối, vì vậy vùng này quyết định đến việc xác định hết mỗi khối.




[18] (FAQ) Đan xen là một cách để tăng tốc độ bộ nhớ. Các câu hỏi về khái niệm đan xen cũng thường xuất hiện trong bài thi vì vậy bạn cần phải hiểu khái niệm này.




Giả sử các đặc tả của băng từ được cho dưới đây, tính toán một cách chính xác số lượng bản ghi có thể được lưu trữ trên một băng từ này[19]

[Đặc trưng của tệp]

| | |
|---|---|
| Chiều dài bản ghi | 80 bytes |
| Hệ số khối (block) | 100 |

[Đặc trưng của băng từ]

| | |
|---|---|
| Mật độ bản ghi | 64 bytes/mm |
| Khoảng cách giữa các khối (IBG) | 15 mm |
| Chiều dài băng từ | 730 m |

**Tính chiều dài khối**

Hệ số khối là 100 → 100 bản ghi có thể được lưu trữ trên một khối
Chiều dài một bản ghi 80 bytes

Do đó tổng số bytes $L_1$ cho mỗi khối, không tính IBG là:

$L_1$ = 80 (bytes/record) ∗ 100 (records/block) = 8,000 (bytes/block).

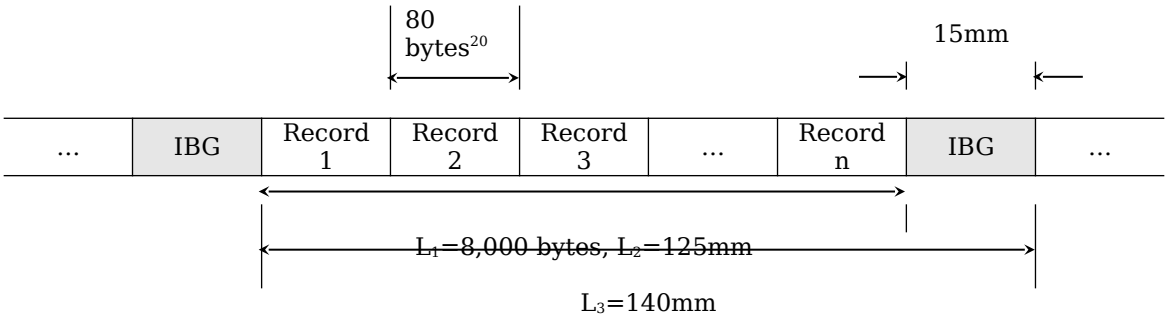
Mật độ bản ghi là 64 bytes/mm → 64 bytes có thể được lưu trữ trên 1 mm của băng từ

Nên chiều dài của một khối $L_2$, không kể IBG là:

$L_2$ = 8,000 (bytes/block)/64 (bytes/mm) = 125 (mm/block).

Do đó chiều dài của một khối $L_3$, kể cả IBG là:

$L_3$ = 125 (mm/block) +15 (mm/block) = 140 (mm/block).




[19] (FAQ) Trong mỗi bài thi, có ít nhất một câu hỏi về tính toán dung lượng hay hiệu năng của băng từ hoặc ổ cứng. Nếu bạn đã biết cách làm, bạn có thể trả lời câu hỏi này bởi vì sự khác biệt chỉ ở số liệu.

[20] (Gợi ý) Bên cạnh "bytes/mm," mật độ bản ghi còn có thể được biểu diễn bằng "cột / mm" hoặc "bpi." Một cột tương tự như 1 byte. "bpi" viết tắt của "bytes / inch," đó là số cột trên inch. Ở đây cần có sự chuyển đổi từ inch sang mm, nhưng bạn cũng không cần phải nhớ công thức chuyển đổi vì công thức sẽ được đưa ra trong bài thi.

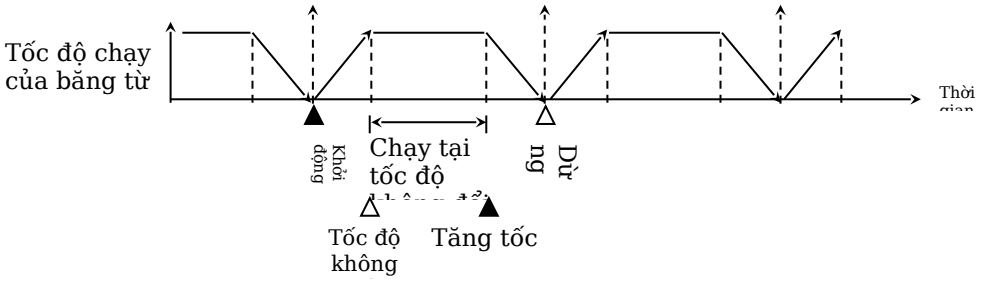
**Tổng số bản ghi có thể được lưu trữ trên một băng từ**

Bây giờ chúng ta sẽ đi tính chính xác tổng số bản ghi có thể được lưu trữ trên băng từ

(1) Tính tổng số khối được có thể được lưu trữ trên băng từ
Vì chiều dài của băng từ là 730 m ($730 * 10^3$ mm), tổng số khối $B_1$ có thể được lưu trên 1 băng từ là:
$B_1$ = (Chiều dài của băng từ) / (chiều dài của một khối)
$= 730 * 10^3 / 140$
$= 5{,}214.285\ldots \fallingdotseq 5{,}214$ (blocks).[21]
Phần phân số 0.285 nhỏ hơn 1 block nên chúng ta bỏ
(2) Tính tổng số bản ghi có thể được lưu trữ trên 1 băng từ
Vì 1 băng từ có thể lưu 5214 khối, và mỗi khối có 100 bản ghi, tổng số bản ghi có thể được lưu trên một băng từ $B_2$ là:
$B_2$ = (Tổng số khối trên băng từ) $*$ (hệ số khối)
$= 5{,}214 * 100$
$= 521{,}400$ (records)

# Tính toán hiệu năng

Tốc độ chạy của băng từ là không đổi khi đọc và ghi dữ liệu. Trên lý thuyết, băng từ bắt đầu gia tốc tại vị trí chính giữa của IBG và bắt đầu đọc hay ghi dữ liệu khi đạt được tốc độ không đổi. Khi bắt gặp một IBG khác, băng từ giảm tốc và dừng lại ở vị trí chính giữa của IBG.




Giả sử các đặc tả của băng từ được cho dưới đây, tính thời gian để băng từ đọc 1 khối (block) dữ liệu[22]

[Đặc tả băng từ]

| | |
|---|---|
| Tốc độ trao đổi dữ liệu | 320 Kbytes/sec |
| Chiều dài bản ghi | 80 bytes |
| Hệ số khối | 100 |
| Thời gian bắt đầu hay dừng lại | 6 milliseconds |

[21] (Gợi ý) Phần thập phân của số khối được loại bỏ ở đây, nhưng trong thực tế nó trở thành một khối ngắn hơn, khối mà có ít bản ghi hơn các khối khác.
[22] (Gợi ý) Trong tính toán hiệu năng, dữ liệu được truyền theo khối, vì vậy chúng ta không cần quan tâm đến chiều dài của IBG.

**Thời gian trao đổi một khối dữ liệu**

Thời gian trao đổi một khối dữ liệu có thể tính được bằng cách chia chiều dài của khối cho tốc độ trao đổi dữ liệu

Thời gian trao đổi dữ liệu của một khối = (Chiều dài một khối) / (tốc độ trao đổi dữ liệu)

Chiều dài khối = (Chiều dài bản ghi) $*$ (hệ số khối)
$= 80 * 100$
$= 8{,}000$ (bytes)

Thời gian trao đổi dữ liệu của một khối = 8,000 (bytes) / 320 (Kbytes/sec)
$= 8{,}000 / (320 * 10^3)$
$= 25 * 10^{-3}$ (seconds) = 25 (milliseconds)

**Thời gian yêu cầu để đọc một khối dữ liệu**

Chúng ta cộng thêm thời gian bắt đầu chạy với thời gian trao đổi dữ liệu của một khối

Thời gian yêu cầu để đọc một khối
= (thời gian bắt đầu) + (thời gian trao đổi dữ liệu của một khối)
= 6 (milliseconds) + 25 (milliseconds) = 31 (milliseconds)

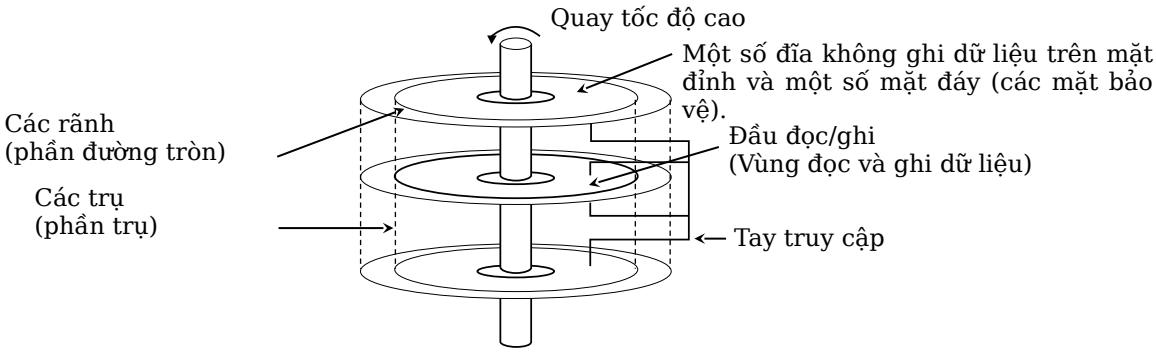
Chú ý rằng thời gian bắt đầu chạy được cộng vào nhưng thời gian dừng lại thì không được cộng. Khi bắt đầu đọc, không có dữ liệu được trao đổi cho đến khi bắt đầu một khối được đọc. Do đó thời gian cho đến khi đạt được điều này là phần thời gian chờ đợi. Sau đó dữ liệu đã được trao đổi, ở thời điểm việc trao đổi dữ liệu hoàn thành hoạt động dừng và xử lý chương trình được thực hiện đồng thời. Vì vậy không cần phải cộng thêm thời gian dừng.[23]

## 2.1.5 Đĩa cứng

| Chủ điểm | ➢ Một đĩa cứng được cấu tạo từ các trụ (cylinder) và rãnh (track)<br>➢ Đĩa cứng loại cung (sector) không có IBGs. |
|---|---|

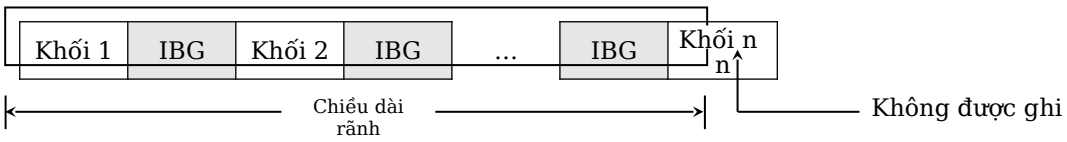
**Đĩa cứng** là một phương tiện truy cập ngẫu nhiên và đọc ghi dữ liệu tốc độ cao, bao gồm 1 đến 10 đĩa tròn được phủ chất liệu từ ở mặt trước và mặt sau của đĩa, quay với tốc độ cao. Nếu chỉ có một đĩa nó được gọi là **đĩa mềm**[24]

[23] (Gợi ý) Thời gian dừng đó thường không được đưa vào như một điều kiện chuẩn trong câu hỏi thi.
[24] (Chú ý) Một phương tiện khác đĩa mềm, có thể đọc và ghi và dễ dàng mang đi mang lại là MO (đĩa quang từ), khá phổ biến bởi vì dung lượng của nó gấp khoảng 600 lần dung lượng của đĩa mềm.


Quay tốc độ cao
Một số đĩa không ghi dữ liệu trên mặt đỉnh và một số mặt đáy (các mặt bảo vệ).
Các rãnh
(phần đường tròn)
Đầu đọc/ghi
(Vùng đọc và ghi dữ liệu)
Các trụ
(phần trụ)
Tay truy cập

# Tính toán dung lượng

Đúng như trên băng từ, dữ liệu được lưu trữ theo khối trên một đĩa cứng. Tuy nhiên, nếu một khối không thể điền đủ vào trong một rãnh, khối đó sẽ không thể được lưu.




Đầu tiên dữ liệu được lưu trữ trên một rãnh và khi rãnh đã được ghi dữ liệu, tiến trình ghi đến rãnh tiếp theo là rãnh nằm trên đường tròn tương ứng của bề mặt tiếp theo. Nói cách khác dữ liệu được lưu trữ trên các đơn vị trụ.

Giả sử rằng đặc tả của một đĩa cứng được cho dưới đây, có bao nhiêu trụ thực sự cần để lưu trữ 100 000 bản ghi?

[Đặc trưng của tệp]

| | |
|---|---|
| Chiều dài bản ghi | 250 bytes |
| Hệ số khối | 8 |

[Đặc tả của đĩa cứng]

| | |
|---|---|
| Số trụ trên đĩa | 400 |
| Số rãnh trên một trụ | 19 |
| Tổng số byte trên một rãnh | 13,000 |
| Khoảng cách giữa các khối | 135 bytes |

**Chiều dài khối (B)**

$$B = (\text{chiều dài bản ghi}) * (\text{hệ số khối}) + (\text{khoảng hở giữa 2 khối})$$
$$= 250 * 8 + 135 = 2{,}135 \text{ (bytes)}$$

**Tổng số khối có thể được lưu trữ trên một rãnh (N)**

N = Chiều dài rãnh / 2,135 = 6.008…≒6 (khối) (Đã cắt bỏ phần lẻ)

Chúng ta lấy chẵn kết quả vì phần thập phân không tạo thành được 1 khối

**Tổng số bản ghi được lưu trữ trên một rãnh ($R_t$)**

$$R_t = N * (\text{hệ số khối})$$
$$= 6 * 8 = 48 \text{ (records)}$$

**Tổng số bản ghi được lưu trữ trên một trụ ($R_s$)**

$$R_s = R_t * (\text{số lượng rãnh trên một trụ})$$
$$= 48 * 19 = 912 \text{ (bản ghi)}$$

**Tổng số trụ cần để lưu trữ 100,000 bản ghi (S)**

S = (tổng số bản ghi) / $R_s$
= 100,000 / 912 = 109.649… ≒110 (trụ) (làm tròn lên).

Nói chung, các thông tin ghi trên đĩa dưới dạng gắn với các đơn vị trụ, vì vậy nếu có một rãnh nào đó không sử dụng trên trụ thì số lượng trụ sẽ được làm tròn đến số nguyên tiếp theo.[25]

# Tính toán hiệu năng

Thời gian truy cập đĩa cứng được tính toán như sau:

Thời gian truy cập = Thời gian đợi + Thời gian trao đổi dữ liệu
= (thời gian tìm kiếm + thời gian trễ) + thời gian trao đổi dữ liệu

Chú ý: Thời gian đợi (thời gian định vị + thời gian trễ) đôi khi cũng được gọi là thời gian truy cập.

**Thời gian tìm kiếm** là thời gian trong đầu đọc/ghi di chuyển đến rãnh nơi đã có dữ liệu được lưu trữ. **Thời gian trễ** là thời gian cho đến khi dữ liệu cần đọc hoặc ghi chuyển đến vị trí đầu đọc. Thời gian tìm kiếm và thời gian trễ được xác định phụ thuộc vào việc đầu đọc đang ở vị trí nào, vì vậy chúng ta sử dụng giá trị trung bình. Trong các câu hỏi ví dụ, thời gian tìm kiếm luôn cho trước, và thời gian trễ có thể tính theo thời gian của một vòng quay, thời gian này nhận được bằng cách ngịch đảo giá trị số vòng quay trên một đơn vị thời gian. Vì thời gian trễ tối thiểu là 0 và tối đa là thời gian một vòng quay nên thời gian trễ trung bình là khoảng thời gian nửa vòng quay.

Giả sử các đặc tả của đĩa cứng được cho sau đây, tính thời gian truy cập thực cho việc đọc dữ liệu bao gồm một khối (5,000 bytes).

[Đặc tả của đĩa cứng][26]

| | |
|---|---|
| Số vòng quay của đĩa cứng: | 2,500 vòng trên phút |
| Dung lượng trên một rãnh: | 20,000 bytes |
| Thời gian truy cập trung bình: | 25 milliseconds |

### Tính toán thời gian trễ trung bình

Số vòng quay của đĩa cứng này là 2,500 vòng trên phút nghĩa là cứ mỗi phút nó quay được 2,500 vòng. Do đó thời gian quay được tính như sau:

Thời gian 1 vòng quay =(1 * 60,000 ms/phút) / 2,500 vòng quay/phút = 24 ms/vòng quay

Cần chú ý các đơn vị của các đại lượng. Tốc độ quay cho là số vòng quay trên phút, trong khi thời gian quay tính trên mili giây. Do đó chúng ta cần

[25] (FAQ) Trong mỗi bài thi, có ít nhất một câu hỏi về tính toán dung lượng hay hiệu năng của băng từ hay đĩa cứng. Nếu bạn đã biết cách làm bạn có thể trả lời câu hỏi này vì sự khác biệt chỉ ở số liệu.

[26] **Seek time/Latency time**: seek time đôi khi còn được gọi là thời gian xác định vị trí. latency time đôi khi còn được gọi là thời gian tìm.

chuyển đổi từ phút sang mili giây (1 phút = 60s = 60,000 ms).

Vì thời gian trễ trung bình bằng ½ thời gian 1 vòng quay nên nó là 12 miliseconds.

**Tính toán tốc độ trao đổi dữ liệu**

Vì một vòng quay cho phép trao đổi dữ liệu trên một rãnh, nên 20,000 bytes được trao đổi trong vòng 24 miliseconds. Do đó 20,000 /24 (bytes/ms) là tốc độ trao đổi dữ liệu. Chúng ta có thể tính thương số này, nhưng vì nó không chia hết nên ta cứ để tỉ số như vậy

### Tính toán thời gian trao đổi dữ liệu

Thời gian yêu cầu để trao đổi 5,000 bytes được tính như sau:

Thời gian trao đổi dữ liệu = (Tổng dữ liệu cần trao đổi) / (Tốc độ trao đổi)
= 5,000 / (20,000 / 24)
= 5,000 / 20,000 ∗ 24
= 6 (msec)

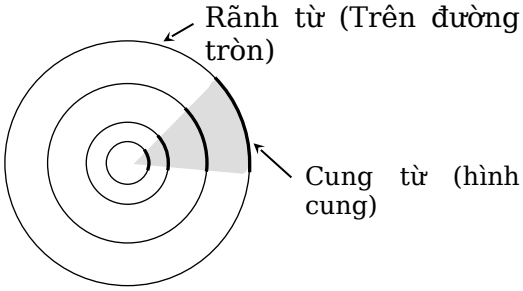
### Tính toán thời gian truy cập

Thời gian truy cập = 25 msec + 12 msec + 6 msec = 43 msec

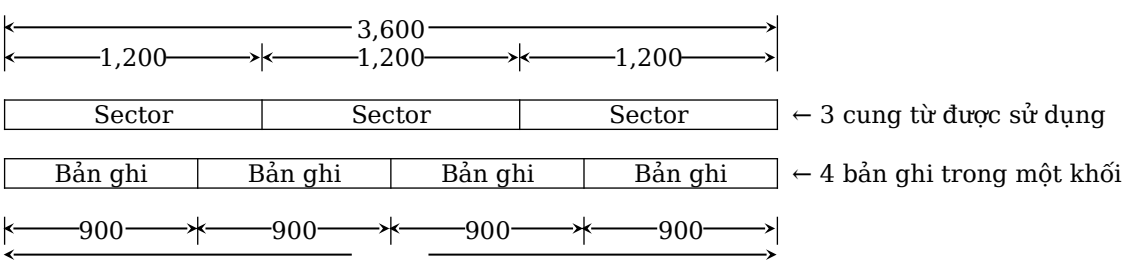
Sự tính toán ở trên là dựa trên tốc độ quay của đĩa cứng. Tuy nhiên, bộ nhớ phụ như đĩa cứng hay băng từ, sự trao đổi dữ liệu với máy tính thông qua các kênh vào ra[27]. Đó là lý do tại sao cần phải cài đặt kênh vào ra có tốc độ phù hợp với tốc độ trao đổi dữ liệu[28]

## Tính toán dung lượng của ổ cứng dựa trên sector

Thuật ngữ sector nói đến cách các phương tiện từ phân vùng trên ổ đĩa mềm hoặc đĩa cứng. Một sector là một cung hình quạt ở trên đĩa chia bởi các tia vẽ từ tâm của các rãnh theo các phần bằng nhau. Vào/ra của phương tiện lưu trữ kiểu sector được thực hiện theo đơn vị sector (từng khối một) và không sử dụng IBG. Mỗi sector có thể được lấp đầy bởi nhiều nhất bản ghi có thể và phần chia còn lại của sector thì không sử dụng.




Ví dụ, giả sử rằng mỗi rãnh từ có 12 cung từ, mỗi cung từ có 1,200 bytes trên một ổ cứng. Để lưu trữ dữ liệu trên đĩa với chiều dài bản ghi là 900 bytes, không có cung từ nào dư ra như hình dưới đây




[27] **Kênh vào/ra**: Đường truyền dữ liệu cho sự trao đổi dữ liệu giữa bộ nhớ phụ và máy tính
[28] (Gợi ý) Tốc độ truyền dữ liệu của ổ cứng được xác định bằng tốc độ quay của đĩa, vì vậy sẽ không có ý nghĩa khi dùng một kênh truyền tốc độ cao. Kênh truyền được lựa chọn theo tốc độ truyền dữ liệu của đĩa cứng.

| 3,600 |

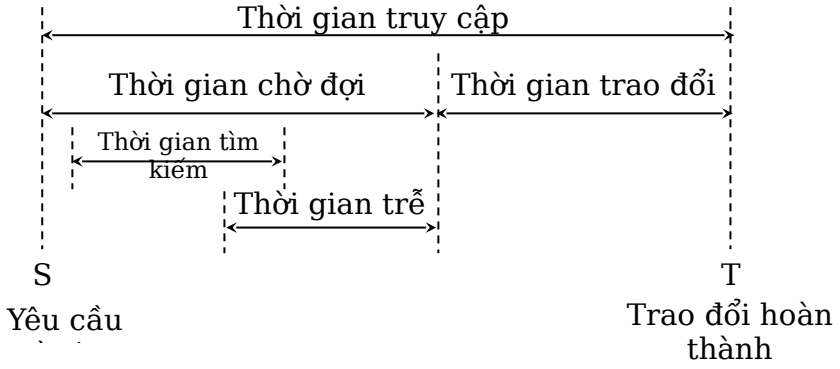
# 2.1.6 Các thuật ngữ liên quan đến hiệu năng/ RAID

| Điểm chính | ➢ Liên quan đến phương tiện lưu trữ có một vài thuật ngữ như thời gian truy cập, thời gian chờ đợi, thời gian trao đổi dữ liệu, thời gian định vị, thời gian trễ.<br>➢ RAID là một hệ thống rất nhiều các đĩa để tăng cường tính tin cậy và tăng tốc độ xử lý. |
|---|---|

Một vài thuật ngữ liên quan đến hiệu năng của các thiết bị lưu trữ bao gồm thời gian truy cập, thời gian chờ đợi, thời gian trao đổi dữ liệu, thời gian định vị, thời gian trễ. RAID, đôi khi còn được gọi là dãy các đĩa là phương pháp điều khiển rất nhiều đĩa làm việc song song như thể chúng chỉ là một đĩa.

## Thuật ngữ liên quan đến hiệu năng

Hình dưới đây mô tả các thời gian xử lý quan hệ với nhau như thế nào, bắt đầu tại thời điểm S khi một khối xử lý yêu cầu vào ra đến đĩa cứng đến thời điểm T khi dữ liệu đã được trao đổi hoàn tất




Bởi vì đầu đọc di chuyển đến rãnh mong muốn trong khi đĩa đang quay nên thời gian quay và thời gian di chuyển chồng lấp lên nhau một phần[29]. Tuy nhiên trong các ví dụ IT, sự chồng lấp này luôn được lờ đi do đó không có vấn đề gì khi định nghĩa thời gian chờ đợi như sau:

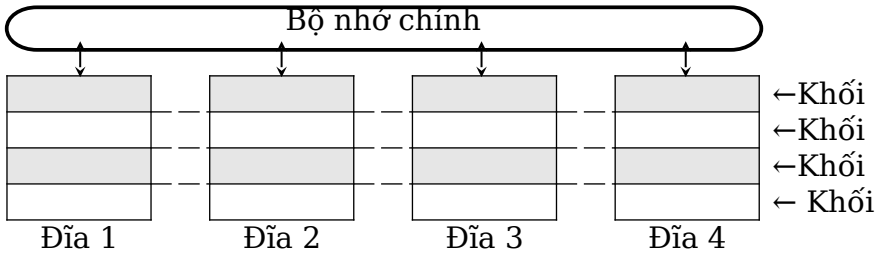
Thời gian chờ đợi = thời gian tìm kiếm + thời gian trễ

## RAID (Redundant Array of Independent Disks)

RAID diễn tả bộ lưu trữ hỗ trợ, trong đó rất nhiều đĩa cứng làm việc song song và được điều khiển như thể là chỉ có 1 đĩa cứng vì vậy tốc độ vào ra được cải thiện và độ tin cậy có thể được nâng cao. Đôi khi thuật ngữ RAID cũng nói đến các bộ lưu trữ phụ hay một phương pháp.

[29] (Gợi ý) Một ổ cứng quay với một đầu từ tiếp xúc với rãnh, vì vậy thời gian định vị và thời gian trễ chồng lấp lên nhau một phần.

RAID cố gắng tăng tốc độ xử lý bằng cách dàn trải các khối dữ liệu trên nhiều đĩa cứng và đọc các khối dữ liệu một cách đồng thời.




RAID có 6 cấp được mô tả dưới đây, từ RAID0 đến RAID5

**RAID0**

Đây là phương pháp ghi các khối dữ liệu với một kích cỡ cố định trên nhiều ổ cứng. Truy cập không bị tập trung vào một đơn vị, vì vậy thời gian vào/ra được rút ngắn.[30]

**RAID1**

Bằng cách lưu trữ dữ liệu giống nhau trên hai đĩa, cấu hình theo kiểu này giúp an toàn dữ liệu.[31]

**RAID2, RAID3, and RAID4**

Cùng với dữ liệu được lưu trữ trên đĩa cứng, có một đĩa được chỉ định là đĩa kiểm lỗi và ngăn ngừa lỗi. RAID 2 có thể sửa lỗi. RAID3 và RAID4 có thể phát hiện lỗi nhưng chúng không thể sửa lỗi. Trong RAID3, dữ liệu được phân vùng theo các bit hoặc bytes trong khi RAID4 phân vùng dữ liệu theo khối.[32]

**RAID5**

Ở đây, mỗi khối dữ liệu được gán cho một giá trị parity (chẵn lẻ). Dữ liệu và parity được ghi trên những đĩa riêng biệt và việc lỗi trên một đĩa có thể được phục hồi.
Dưới đây là ví dụ của RAID5. Ở đây 4 đĩa được nhóm trong 1 nhóm

| Đĩa 1 | Đĩa 2 | Đĩa 3 | Đĩa 4 |
|---|---|---|---|
| Khối 1 | Khối 2 | Khối 3 | Parity[33] 1~3 |
| Khối 4 | Khối 5 | Parity 4~6 | Khối 6 |
| Khối 7 | Parity 7~9 | Khối 8 | Khối 9 |
| Parity 10~12 | Khối 10 | Khối 11 | Khối 12 |

Ở đây, dữ liệu được chia thành các khối với độ dài xác định, và các khối này được hình thành một đơn vị. Ví dụ, khối 1 đến khối 3 là một đơn vị và với

[30] (Gợi ý) RAID0 chỉ có tính năng là vào ra phân tán, vì vậy nó không được dùng để cải thiện độ tin cậy.
[31] (Chú ý) RAID1 được gọi là soi gương bởi vì dữ liệu giống nhau được lưu trên các đĩa cứng riêng biệt.
[32] (Chú ý) Còn có RAID0+1, kết hợp của RAID0 và RAID1 đã được sử dụng.
[33] (Chú ý) Tính chẵn lẻ của RAID5 sử dụng tổng logic loại trừ của nhiều khối (XOR). Vì thế ngay cả khi một đĩa cứng bị hỏng, dữ liệu bị mất vẫn có thể được phục hồi bằng cách lấy tổng loại trừ của các khối khác.

mỗi bit giá trị tổng logic loại trừ từ khối 1 đến khối 3 được ghi trên một đĩa riêng biệt là các giá trị chẵn lẻ từ 1 đến 3. Tương tự, giá trị tổng logic loại trừ từ khối 4 đến khối 6 được ghi trên đĩa riêng biệt là giá trị chẵn lẻ từ 4 đến 6.

Ngoài ra còn có RAID6, trong đó các giá trị chẵn lẻ được lưu riêng biệt giống như RAID5 và dữ liệu có thể được phục hồi ngay cả khi hai đĩa bị lỗi.[34]

## 2.1.7 Thiết bị lưu trữ phụ trợ/Các đơn vị vào ra

| **Chủ điểm** | ➢ Thiết bị lưu trữ phụ bao gồm đĩa cứng, băng từ, đĩa quang từ, CDs, DVDs…<br>➢ Các đơn vị vào ra bao gồm bàn phím, máy quét ảnh, màn hình, máy in… |
|---|---|

Tất cả các thiết bị lưu trữ ngoại trừ bộ nhớ chính được gọi là thiết bị lưu trữ phụ (hỗ trợ). Thiết bị lưu trữ phụ được dùng để bù đắp sự thiếu dung lượng của bộ nhớ chính. Nhìn chung, thiết bị lưu trữ phụ có dung lượng lớn hơn dung lượng của bộ nhớ chính.

**Các khối vào ra** bao gồm các khối vào – nơi dữ liệu được đưa vào máy tính và các khối ra nơi dữ liệu được đưa từ máy tính ra ngoài. Một khối được trang bị cả hai chức năng vào và ra thì được gọi là khối vào/ra.

### Thiết bị lưu trữ phụ

Các thiết bị lưu trữ phụ thông dụng bao gồm các phương tiện sau. Trong quá khứ đã có một thời gian khi băng từ và ổ mềm là các phương tiện phổ biến; tuy nhiên gần đây các phương tiện chủ yếu gồm đĩa cứng, đĩa quang từ, CDs và DVDs.[35]

| Phương tiện | | Dung lượng | Ghi lại được | Thuộc tính |
|---|---|---|---|---|
| Đĩa cứng | | Nhỏ hơn 300GB | Có | Mostly built-in |
| DVD | DVD-ROM | 4.7 to 9.4BG | Không | Thay thế cho CD-ROM |
| | DVD-R<br>DVD+R | 3.95 to 7.9GB | Có, chỉ 1 lần | Có thể chơi được trên DVD-ROM |
| | DVD-RAM<br>DVD-RW<br>DVD+RW | 3.95 to 7.9GB | Có, nhiều lần | Rất nhiều đặc tính |
| CD | CD-ROM | 700MB | Không | Cho các gói phần mềm. |
| | CD-R | 700MB | Có, chỉ 1 lần | Cho việc sao lưu. |

[34] (FAQ) Thường xuyên trong các câu hỏi có dạng: “Câu nào trong các câu sau đây là phù hợp liên quan đến RAID…?” Hãy nhớ sự khác nhau giữa RAID0 và RAID1.
[35] (Chú ý) DVD cũng là đĩa quang giống như CD, nhưng bằng việc giảm bước sóng của tia laze, DVD có dung lượng lưu trữ lớn hơn. Mật độ các bản ghi trên đĩa DVD cũng lớn hơn.

| | | | | |
|---|---|---|---|---|
| | CD-RW | 700MB | Có | Cần các ổ chuyên dụng cho việc ghi lại |
| Đĩa quang từ[36] (MO) | | 128, 230, 640MB, 1.3GB | Có | Được ghi bằng từ tính và ánh sáng đọc bằng ánh sáng |
| Đĩa mềm | | 1.4MB | Có | Thuận tiện cho việc mang đi mang lại |
| Đĩa cứng | | Một vài GB | Có | Dùng cho sao lưu |
| DAT[37] | | Lớn nhất vào khoảng 24GB | Yes | Dùng cho sao lưu |

[36] (Gợi ý) Một đĩa quang từ sử dụng ánh sáng và từ tính để ghi dữ liệu nhưng chỉ dùng ánh sáng để đọc dữ liệu.
[37] (Chú ý) DAT là một đơn vị lưu trữ âm thanh trên một băng từ sử dụng tín hiệu số. Ban đầu nó được thiết kế cho âm nhạc nhưng ngày nay nó được dùng như một hệ thống sao lưu vì giá thành rẻ.

# Các thiết bị vào/ra

Các thiết bị vào gồm có bàn phím, máy quét ảnh, bảng, các thiết bị điểm – con trỏ. Các thiết bị ra bao gồm màn hình, máy in…

### Các thiết bị vào

Thiết bị vào thông dụng nhất là bàn phím và các thiết bị con trỏ. Bàn phím được sử dụng để nhập các con số và kí tự trong khi các thiết bị con trỏ được sư dụng để nhập vào các giá trị tọa độ. Các thiết bị vào khác được chỉ ra trong bảng dưới đây.

| Thiết bị | Chức năng |
|---|---|
| OCR | Thiết bị này đọc các ký tự viết tay và các kí tự in bằng quang học |
| OMR | Thiết bị này đọc các dấu viết tay bằng quang học |
| Chuột | Thiết bị này dùng để nhập tọa độ vị của con trỏ chuột |
| Tablet | Thiết bị đặc biệt như bút quang được sử dụng để nhập vào vị trí tọa độ |
| Đọc mã vạch | Đọc barcodes (mã vạch) |
| Máy quét ảnh | Đọc dữ liệu ảnh như các bức vẽ, ảnh chụp chuyển chúng sang dữ liệu số |
| | |

### Các thiết bị ra

Thiết bị ra phổ biến nhất là màn hình và máy in.[38]

| Thiết bị | | Mô tả |
|---|---|---|
| Màn hình | CRT | Thiết bị này sử dụng ống phóng catot, không đắt và thường là những màn hình rộng |
| | Tinh thể lỏng[39] | Sử dụng tinh thể lỏng, giá thành đắt nhưng mỏng và tiết kiệm năng lượng. |
| Printer | Máy in laser | Về cơ bản giống như máy copy, chất lượng in tốt nhưng giá thành đắt. |
| | Máy ín phun | In bằng cách phun mực, nhỏ và giá thành không đắt. |
| | Máy in nhiệt | Chất lượng in tốt nhưng yêu cầu giấy in đặc biệt. |
| | Máy in truyền nhiệt | Nung chảy mực, chất lượng in tốt, nhưng giá hoạt động đắt |
| | Máy in điểm | Gây ồn nhưng giá thành rẻ có thể in hai mặt |
| | Plotter | Máy in cho các bản vẽ thiết kế. |

[38] (Chú ý) **OLED**: là một loại màn hình sử dụng công nghệ diot phát quang ánh sáng hữu cơ. Nó sử dụng các vật liệu hữu cơ phát ra ánh sáng khi có điện thế đặt vào. So sánh với LCDs, nó có góc nhìn rộng hơn, độ tương phản tốt hơn, và tốc độ phản hồi cao hơn, thêm vào đó nó còn mỏng hơn và sáng hơn.

[39] (Note) **Màn hình tinh thể lỏng (LCD)**: Màn hình tinh thể lỏng có thể có nhiều loại như TFT và STN (ngày nay thì dòng chính là DSTN). STN có một cấu trúc đơn giản với giá thành sản xuất thấp, tuy nhiên độ phân giải và độ tương phản cũng thấp. DSTN là phiên bản cải thiện của STN, với độ tương phản được tăng lên. TFT có độ tương phản và độ phân giải tương đương với CRT nhưng giá thành đắt hơn.

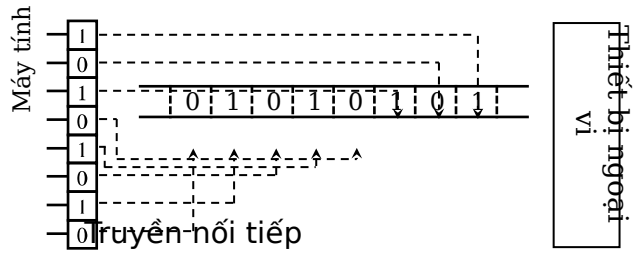
# 2.1.8 Các giao tiếp vào và ra

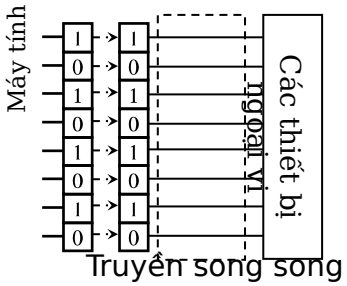
| Điểm chính | ➢ **Giao tiếp vào và ra bao gồm SCSI, USB…**<br>➢ **USB trang bị chức năng cắm nóng và "cắm và chạy"** |
|---|---|

Giao tiếp vào và ra là các giao tiếp cho phép kết nối thiết bị ngoại vi như máy in, đĩa cứng với PC và cho việc truyền nhận dữ liệu. Tùy thuộc vào từng loại, việc truyền nhận có thể là nối tiếp hoặc song song.[40]

## Các phương pháp truyền dữ liệu

Có hai phương pháp truyền dữ liệu giữa máy tính và thiết bị ngoại vi: truyền nối tiếp và truyền song song. **Truyền nối tiếp** là phương pháp truyền mà dữ liệu ra từ máy tính là nối tiếp, 1 bit ở một thời điểm. **Truyền song song** là phương pháp truyền mà các bit dữ liệu được truyền một cách song song. Ví dụ nếu truyền 8 bit song song thì sẽ có 8 đường tín hiệu.




[40] (FAQ) Trong các câu hỏi thi thường có các câu quan tâm đến sự kết hợp giữa giao tiếp vào ra và phương pháp truyền dữ liệu. Cần phải biết các giao tiếp vào ra chung và các phương pháp truyền dữ liệu.

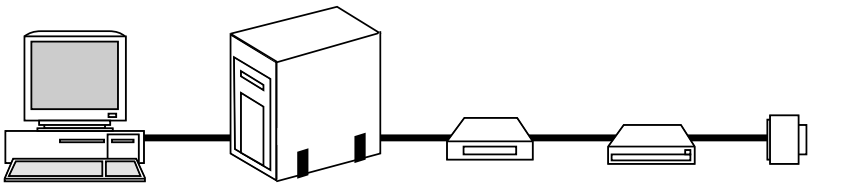
# Các loại giao tiếp vào ra

Dưới đây là bảng tổng hợp các giao tiếp vào ra phổ biến được dùng hiện nay. USB là giao tiếp được dùng phổ biến nhất hiện nay, được hỗ trợ các chức năng cắm và chạy (plug and play[41]) và cắm nóng (hot plug) [42]

| Loại | Phương pháp truyền | Thuộc tính |
|---|---|---|
| RS-232C | Nối tiếp | Ban đầu được sử dụng để nối máy tính PC với modem, ngày nay cũng được dùng cho các đơn vị vào ra |
| SCSI | Song song | Daisy chain cho phép kết nối với các đơn vị tốc độ cao |
| Centronics | Song song | Kết nối máy tính với máy in, tốc độ cao nhưng không cho phép khoảng cách dài |
| GPIB | Song song | Kết nối máy tính với các thiết bị đo lường, được biết đến như là IEEE-488 |
| USB | Nối tiếp | Kết nối tới 127 đơn vị trong cấu hình dạng cây |
| IEEE1394 [43] | Nối tiếp | Kết nối lên tới 63 đơn vị trong daisy chain hoặc cấu hình dạng cây. |

### Daisy chain

Đây là phương pháp kết nối sử dụng trong SCSI và GPIB, các thiết bị ngoại vi được kết nối dọc theo một dường dây. Thiết bị cuối cùng trên đường dây yêu cầu một điện trở kết thúc được gọi là thiết bị cuối.




### USB

USB có hai chế độ: Chế độ tốc độ cao 12Mb/s và chế độ tốc độ thấp 1.5Mb/s. Trong chế độ tốc độ cao, các đơn vị tốc độ cao như máy in hay máy quét được kết nối. Trong chế độ tốc độ thấp, các đơn vị tốc độ thấp như bàn phím hay chuột được kết nối. Ngày nay, USB 2.0 đã tăng chế độ tốc độ cao của nó lên đến 480Mb/s, vì vậy hầu hết các đơn vị ngoại vi có thể được kết nối.

[41] **Cắm-và-chạy (Cắm và chạy)**: Điều này nói đến chức năng tự động cài đặt và thiết lập driver của thiết bị khi các đơn vị ngoại vi hoặc các card mở rộng kết nối với máy tính. Hệ điều hành kiểm tra tất cả các đơn vị đã kết nối với máy tính khi khởi động, cài đặt các driver thiết bị yêu cầu. Nếu hệ điều hành không có driver thiết bị của đơn vị trong thư viện của nó, nó sẽ yêu cầu cài đặt driver cho thiết bị và nếu cần thiết sẽ khởi động lại máy tính một cách tự động.

[42] **Cắm nóng**: Đây là chức năng cho phép chức năng cắm và chạy trong khi nguồn của máy tính và thiết bị ngoại vi đang bật.

[43] **IEEE 1394**: Chuẩn tốc độ truyền là 100Mb/s, 200Mb/s, hoặc 400Mb/s. Chuẩn này được trang bị chức năng trao đổi nóng (thiết bị ngoại vi có thể kết nối hoặc ngắt kết nối mà không cần phải tắt nguồn).

### Các giao tiếp khác

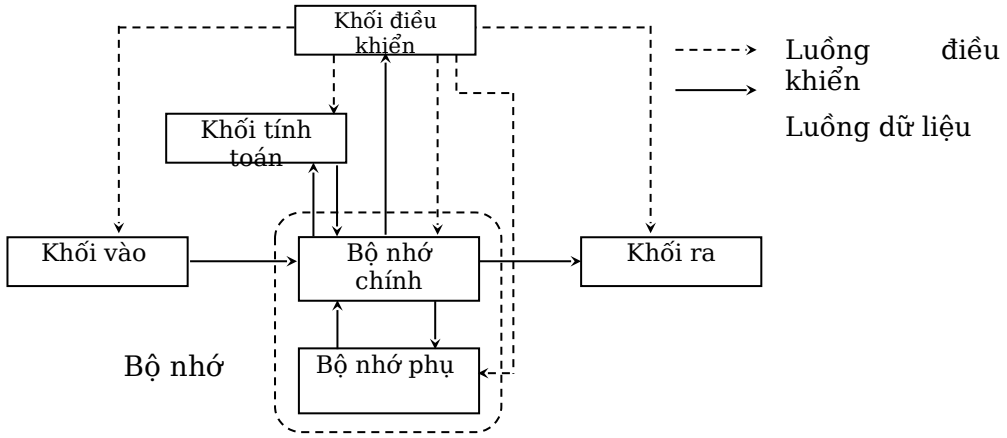
Ngoài các giao tiếp trên, còn có các giao tiếp khác như IDE (kết nối với đĩa cứng), ATA (IDE được chuẩn hóa bởi ANSI) ATAPI (kết nối ATA với các đơn vị khác ổ cứng như ổ CDROM…) Tuy nhiên các thiết bị ổ đọc CD và ghi CD ngày nay thường kết nối thông qua SCSI và USB.

## Câu hỏi nhanh

**Q1** So sánh SRAM và DRAM:

| Các mục so sánh | DRAM | SRAM |
|---|---|---|
| Làm tươi | | |
| Mức độ tích hợp | | |
| Tốc độ truy cập | | |
| Giá đơn vị trên bit | | |
| Cách sử dụng | | |

**Q2** Trong các thành phần tạo nên máy tính, những thành phần nào tạo nên bộ xử lý trung tâm (CPU)?




**Q3** Giải thích vai trò của bộ nhớ đệm.

**Q4** Mô tả các đặc điểm của RAID.

**Q5** Giao tiếp USB là nối tiếp hay song song?

**A1**

| Mục so sánh | DRAM | SRAM |
|---|---|---|
| Làm tươi | Cần thiết | Không cần |
| Mức độ tích hợp | Cao (Dung lượng lớn) | Thấp (dung lượng nhỏ) |
| Tốc độ truy cập | Chậm | Nhanh |
| Giá đơn vị trên bit | Rẻ | Đắt |
| Sử dụng | Bộ nhớ chính | Bộ nhớ đệm |

**A2** Tập hợp của đơn vị điều khiển và đơn vị xử lý

**A3** Bộ nhớ tốc độ cao dung lượng nhỏ nằm giữa CPU hoặc thanh ghi và bộ nhớ chính. Tốc độ truy cập bộ nhớ chính là chậm hơn so với CPU hoặc thanh ghi, vì thế để tăng hiệu quả xử lý của CPU bằng cách đưa các dữ liệu truy cập thường xuyên vào bộ nhớ đệm

**A4** Đây là thiết bị lưu trữ phụ mà rất nhiều ổ cứng được nối song song và được điều khiển như thể là chúng chỉ là một ổ đĩa. Vì vậy tốc độ vào ra được cải thiện và độ tin cậy được nâng cao. Đôi khi thuật ngữ RAID muốn nói đến phương pháp thay vì việc lưu trữ. RAID cố gắng đẩy nhanh tốc độ xử lý bằng cách dàn trải các bản ghi trên nhiều đĩa và đọc các bản ghi một cách đồng thời.

**A5** Giao tiếp nối tiếp (“USB” viết tắt của “Universal Serial Bus.”)

# 2.2 Hệ điều hành

## Giới thiệu

Máy tính không thể hoạt động nếu chỉ có phần cứng. Chúng chỉ có đầy đủ chức năng bằng cách sử dụng một phần mềm gọi là hệ điều hành (OS).

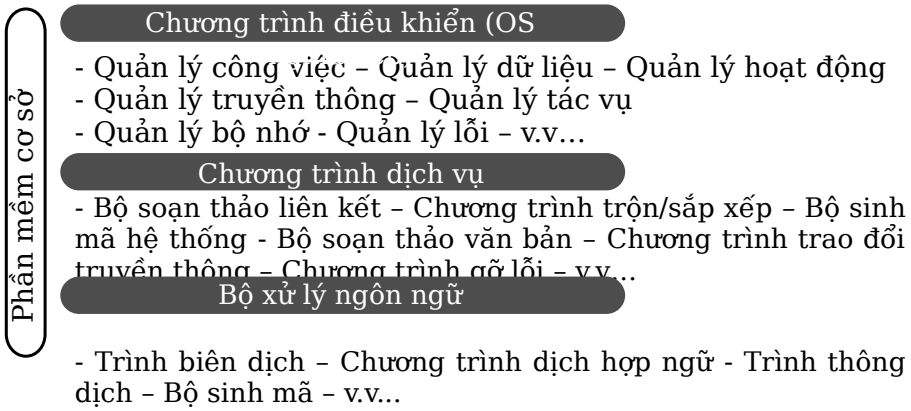
## 2.2.1 Cấu hình và các mục tiêu của hệ điều hành

| Điểm chính | ➢ Định nghĩa khái quát và định nghĩa cụ thể hệ điều hành<br>➢ Mục tiêu của hệ điều hành là sử dụng hiệu quả các máy tính. |
|---|---|

Không có định nghĩa rõ ràng cho hệ điều hành. Một phần mềm cơ sở được gọi là một hệ điều hành trong một ngữ cảnh rộng trong khi một chương trình điều khiển được gọi là một hệ điều hành trong một ngữ cảnh hẹp.

### Cấu hình của hệ điều hành

Hệ điều hành là một phần mềm cơ sở mà điều khiển hoàn toàn và quản lý toàn bộ các hoạt động của phần cứng và phần mềm của một hệ thống máy tính. Một chương trình được coi như phần mềm cơ sở và vai trò của nó được chỉ ra như dưới đây [44].


Phần mềm cơ sở
Chương trình điều khiển (OS
- Quản lý công việc – Quản lý dữ liệu – Quản lý hoạt động
- Quản lý truyền thông – Quản lý tác vụ
- Quản lý bộ nhớ - Quản lý lỗi – v.v...
Chương trình dịch vụ
- Bộ soạn thảo liên kết – Chương trình trộn/sắp xếp – Bộ sinh mã hệ thống - Bộ soạn thảo văn bản – Chương trình trao đổi truyền thông – Chương trình gỡ lỗi – v.v...
Bộ xử lý ngôn ngữ
- Trình biên dịch – Chương trình dịch hợp ngữ - Trình thông dịch – Bộ sinh mã – v.v...

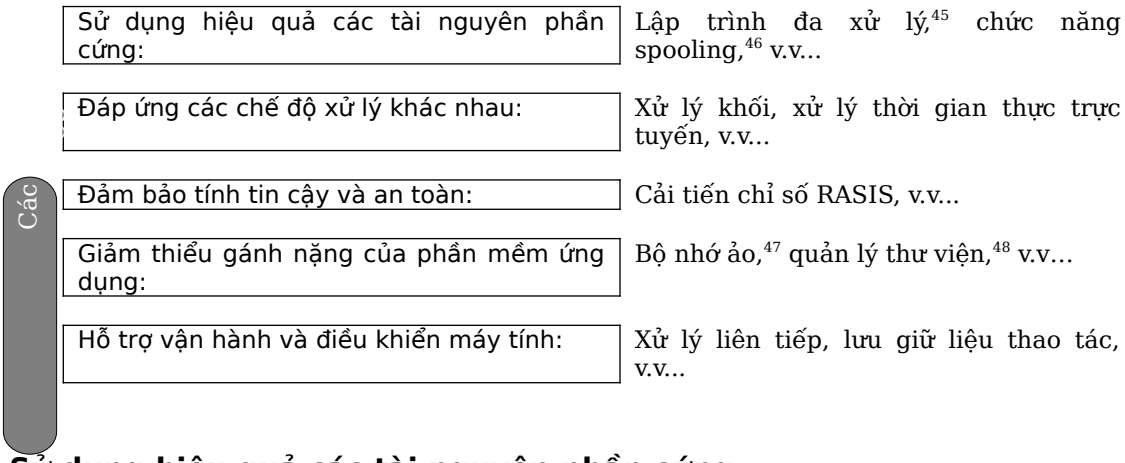
Các chương trình dịch vụ và các bộ xử lý ngôn ngữ đôi khi cũng được gọi là

[44] (Gợi ý) Các hệ điều hành trên máy tính cá nhân bao gồm Windows XP, Mac OS X, và OS/2. Đối với các máy trạm bao gồm Windows Server 2003, UNIX, và Mac OS X Server. Đối với các máy mục đích chung, thường là MVS phát triển bởi IBM. Ngoài ra còn có, chương trình hệ điều hành miễn phí Linux tương thích với UNIX.

các chương trình xử lý, chạy trên các chương trình điều khiển. Vì lý do này, chương trình điều khiển được gọi là một hệ điều hành theo nghĩa hẹp.

# Mục đích của hệ điều hành

Hệ điều hành cố gắng cải tiến hiệu năng của toàn bộ hệ thống bằng cách loại bỏ các hoạt động không cần thiết, bỏ qua các tài nguyên ngoại vi của máy tính và vận hành hệ thống máy tính một cách hiệu quả. Mục đích của hệ điều hành được chỉ ra như trong hình dưới đây.


Các
Sử dụng hiệu quả các tài nguyên phần cứng:
Lập trình đa xử lý,[45] chức năng spooling,[46] v.v...
Đáp ứng các chế độ xử lý khác nhau:
Xử lý khối, xử lý thời gian thực trực tuyến, v.v...
Đảm bảo tính tin cậy và an toàn:
Cải tiến chỉ số RASIS, v.v...
Giảm thiểu gánh nặng của phần mềm ứng dụng:
Bộ nhớ ảo,[47] quản lý thư viện,[48] v.v...
Hỗ trợ vận hành và điều khiển máy tính:
Xử lý liên tiếp, lưu giữ liệu thao tác, v.v...

### Sử dụng hiệu quả các tài nguyên phần cứng

Các tài nguyên phần cứng bao gồm khối xử lý trung tâm (CPU), bộ nhớ, các khối vào ra (I/O) (bao gồm các kênh vào ra), v.v...Hệ điều hành điều khiển các nguồn tài nguyên để chúng có thể được sử dụng một cách hiệu quả.

### Đáp ứng các chế độ xử lý khác nhau

Một máy tính có thể quản lý các chế độ xử lý khác nhau như xử lý theo khối, xử lý theo khối từ xa, xử lý trực tuyến, xử lý thời gian thực, và xử lý chế độ tương tác. Đặc biệt, khi xử lý trực tuyến trở nên phổ biến thì phạm vi của các ứng dụng máy tính càng được mở rộng nhanh chóng.

### Tính tin cậy và an toàn

Thước đo của tính tin cậy và an toàn được đưa ra bởi thuật ngữ RASIS. Nó là viết tắt chữ cái đầu của các từ: tính tin cậy (Reliability), tính hiệu lực (Availability), khả năng phục vụ (Serviceability), tính toàn vẹn (Integrity) và

[45] **Lập trình đa xử lý**: Là một cơ chế trong đó các chương trình được xử lý qua lại trên một CPU để nó có thể xuất hiện mặc dù nhiều chương trình đang hoạt động cùng lúc.
[46] **Spooling**: Là một kỹ thuật sử dụng các đĩa cứng tốc độ cao như một khối vào/ra ảo. Ví dụ, việc in trực tiếp trên một máy in tốc độ thấp sẽ làm chậm tốc độ xử lý. Thay vào đó, các kết quả xuất ra có thể được ghi vào một đĩa cứng tốc độ cao trước, và sau đó một chương trình phục vụ, chỉ quan tâm đến xuất dữ liệu, thực hiện việc in khi CPU không bận.
[47] **Bộ nhớ ảo**: Đây là một kỹ thuật để mở rộng dung lượng của bộ nhớ chính để các chương trình lớn có thể được nạp vào trong bộ nhớ chỉ một lần. Thông thường, bộ nhớ ngoài như là đĩa cứng được sử dụng như bộ nhớ ảo.
[48] **Quản lý thư viện**: Đây là một chức năng mà tích lũy một cách hệ thống các chương trình thô sơ, các chương trình bổ trợ, các modules tải, và các chương trình khác được phát triển. Cho phép quản lý tích hợp các tài nguyên phần mềm mà được quản lý riêng lẻ.

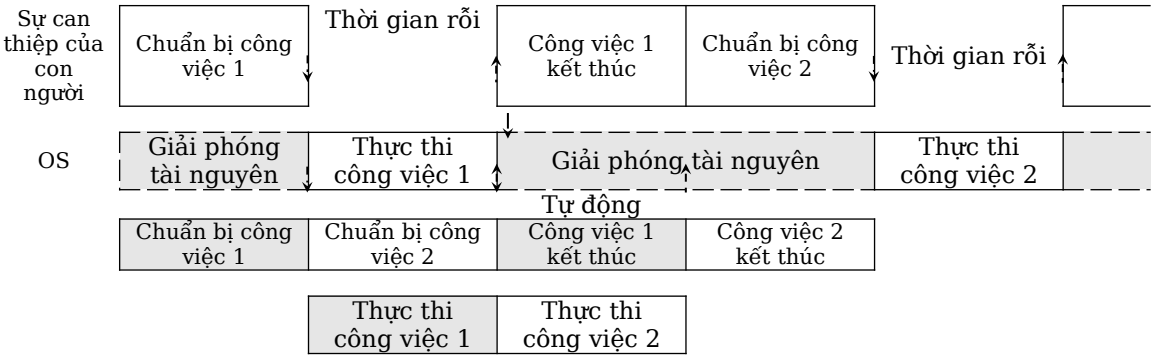
sự an toàn (Security).

### Giảm thiểu gánh nặng của phần mềm ứng dụng

Các phần mềm ứng dụng được biết như là một chương trình chạy dưới điều khiển của hệ điều hành. Hệ điều hành cung cấp một môi trường cho các chương trình ứng dụng có thể được thực thi một cách hiệu quả.

### Hỗ trợ vận hành và điều khiển máy tính

Hệ điều hành loại bỏ sự can thiệp của con người nhiều nhất có thể, bởi vì nó xử lý các chương trình liên tục và ghi lại trạng thái hoạt động . Thông tin các trạng thái hoạt động đã ghi lại sẽ được sử dụng để kiểm tra các tình huống lỗi xảy ra.[49]

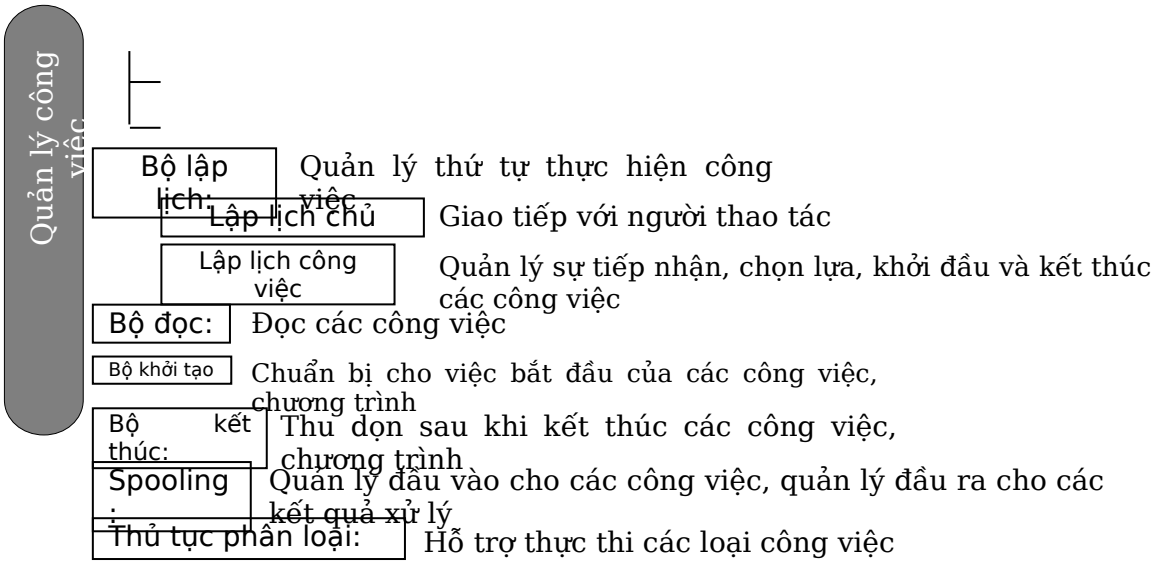

Sự can thiệp của con người
Chuẩn bị công việc 1
Thời gian rỗi
Công việc 1 kết thúc
Chuẩn bị công việc 2
Thời gian rỗi
OS
Giải phóng tài nguyên
Thực thi công việc 1
Giải phóng tài nguyên
Thực thi công việc 2
Tự động
Chuẩn bị công việc 1
Chuẩn bị công việc 2
Công việc 1 kết thúc
Công việc 2 kết thúc
Thực thi công việc 1
Thực thi công việc 2

## 2.2.2 Quản lí công việc

| **Điểm chính** | ➢ Công việc là các đơn vị tác vụ được đưa ra trong máy tính, bao gồm nhiều chương trình (các bước công việc).<br>➢ Quản lí công việc có các chức năng như lập lịch và kỹ thuật spooling. |
|---|---|

Một trong những chức năng của hệ điều hành là "quản lí công việc". Trong quản lí công việc, quyền ưu tiên của các công việc sẽ được xác định, và các công việc được đồng bộ với nhau. Trong xử lý theo khối, hệ điều hành phân tích các nội dung của ngôn ngữ điều khiển công việc (JCL – Job Control Language) để đăng kí tài nguyên[50] và lập lịch cho các công việc. Trong chế độ xử lí tương tác, hệ điều hành phân tích các lệnh đầu vào để đăng kí tài nguyên và lập lịch thực hiện. Ngoài ra, quản lí công việc có các chức năng khác như điều độ công việc và phân loại các thủ tục.

[49] (FAQ) Liên quan đến hệ điều hành, có nhiều câu hỏi liên hệ đến kiến thức các thuật ngữ. Cần đảm bảo hiểu các thuật ngữ như multiprogramming, virtual memory, và spooling function.
[50] **Resource**: Tài nguyên là một thiết bị hoặc một khối thuộc rất nhiều loại khác nhau cần thiết cho hoạt động của máy tính. Đó là bất cứ các thiết bị gì liên quan đến bộ nhớ, vào, ra, điều khiển và các chức năng khác; đặc biệt bao gồm CPU, bộ nhớ chính, và các tệp.


Quản lý công việc
Bộ lập lịch:
Quản lý thứ tự thực hiện công việc
Lập lịch chủ
Giao tiếp với người thao tác
Lập lịch công việc
Quản lý sự tiếp nhận, chọn lựa, khởi đầu và kết thúc các công việc
Bộ đọc:
Đọc các công việc
Bộ khởi tạo
Chuẩn bị cho việc bắt đầu của các công việc, chương trình
Bộ kết thúc:
Thu dọn sau khi kết thúc các công việc, chương trình
Spooling :
Quản lý đầu vào cho các công việc, quản lý đầu ra cho các kết quả xử lý
Thủ tục phân loại:
Hỗ trợ thực thi các loại công việc

## Bộ lập lịch

Trong quản lí công việc, các công việc được thực thi liên tiếp theo một lịch chủ và lịch công việc. Lịch chủ đóng vai trò giao tiếp với người thao tác thông qua bảng điều khiển.[51] Lịch công việc quản lý sự tiếp nhận, lựa chọn, khởi đầu và kết thúc của các công việc.

## Bộ đọc

Đọc nội dung của JCL, phân tích chúng, lập lịch cho các công việc và đặt chúng vào một hàng đợi.

## Bộ khởi tạo

Chọn lựa các chương trình có quyền thực thi cao trong số các chương trình trong hàng đợi và đăng kí các tài nguyên cần thiết cho các chương trình đó.

## Bộ kết thúc

Giải phóng các tài nguyên mà chương trình đã sử dụng xong. Nếu có một chương trình khác tiếp theo, bộ kết thúc sẽ tiếp tục khởi động bộ khởi tạo.

## Kỹ thuật Spooling (Spool)

Spooling là chức năng vào/ra của các công việc độc lập của các chương trình. Các kết quả xuất đến các khối tốc độ thấp như máy in, đầu tiên được lưu trữ trong một tệp spool. Sau đó, khi chương trình kết thúc, kết quả đầu ra được in trên máy in từ tệp spool bởi chương trình dịch vụ của hệ điều hành[52].

Lý do sử dụng kỹ thuật này là khi khối vào ra chậm, việc thực hiện xử lý vào ra trực tiếp sẽ làm giảm tốc độ xử lý của máy tính.[53]

## Thủ tục phân loại

Trong điều khiển thực thi công việc, các xử lý thông thường như dịch ngôn ngữ được thực hiện như sau: Một tập hợp JCLs được đăng kí cùng nhau tại một vị trí riêng biệt, và đăng kí một tập các JCL được gọi để thực hiện các chương trình. Bằng cách này, máy tính ngăn chặn được các lỗi JCL. Tập hợp các JCL được đăng kí gọi là phân loại các thủ tục.

---

[51] **Console panel (Bảng điều khiển)**: Là khối thực hiện các thao tác tương tác với hệ thống máy tính thông qua các phím điều khiển và giám sát các hoạt động của nó ở nơi hệ thống truyền thông bị hỏng hóc, v.v… Nó bao gồm một bàn phím và một màn hình hiển thị.

[52] (Gợi ý) Các tệp spool thường sử dụng trong một đĩa cứng. Chương trình dịch vụ thực hiện gửi các tệp spool đến một máy in thường được gọi là một chương trình thư ký (output writer).

[53] (FAQ) Các câu hỏi thi phần quản lý công việc thường là về kỹ thuật spooling. Chúng thường ở dạng lựa chọn thuật ngữ đúng, vì vậy cần phải hiểu chính xác kỹ thuật spooling.

# 2.2.3 Quản lí tác vụ

| Điểm chính | ➢ Tác vụ là một đơn vị thực thi nhỏ nhất sử dụng một tài nguyên.<br>➢ Sử dụng cơ chế ngắt để chuyển đổi giữa các tác vụ. |
|---|---|

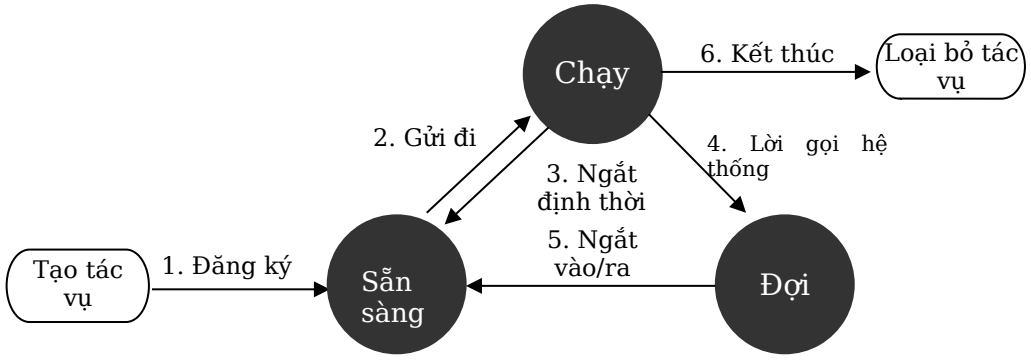
Một trong những chức năng của hệ điều hành là "quản lí tác vụ". Quản lí tác vụ là chức năng điều khiển việc thực thi các chương trình và bao gồm các thủ tục khác nhau như đồng bộ hóa điều khiển chương trình, cấp phát động tài nguyên để thực thi chương trình, và quản lí quyền ưu tiên thực thi của các chương trình. Nó cũng kiểm soát các loại điều khiển ngắt khác nhau.

## Tác vụ và công việc

Công việc (jobs) là một đơn vị xử lý theo ngữ cảnh người sử dụng (một chương trình đơn lẻ hoặc một tập các chương trình được thực thi liên tiếp). Trái lại, tác vụ (tasks) là một đơn vị xử lý bên trong được thực thi dưới hệ điều hành. Tác vụ được xem như là một đơn vị xử lý mà phân chia từ một tiến trình chương trình.[54]

## Điều khiển thực thi tác vụ

Quản lí tác vụ tạo ra các tác vụ được yêu cầu đáp ứng một lệnh và giám sát quá trình thực thi. Khi một tác vụ đã tạo ra không còn cần thiết nữa, nó sẽ được loại bỏ. Biểu đồ dịch chuyển trạng thái của các tác vụ được chỉ ra như sau đây:




| |
|---|
| 1. Một tác vụ được tạo ra sẽ được đặt vào trong hàng đợi tác vụ → chuyển tới trạng thái ready (sẵn sàng). |
| 2. Để thực thi, tác vụ được chuyển tới trạng thái Running (chạy) thông qua tác vụ Dispatcher (gửi) → chuyển tới trạng thái running. |
| 3. Hết thời gian cho phép, tác vụ rút khỏi trạng thái Running để nhường cho một tác vụ khác có quyền ưu tiên cao hơn → chuyển đến trạng thái Ready. |
| 4. Rút khỏi một lệnh vào/ra → chuyển đến trạng thái Waiting (đợi). |
| 5. Trạng thái đợi được thực hiện lần nữa sau khi hoàn thành vào/ra → chuyển đến trạng thái sẵn sàng. |

[54] **Process**: (Tiến trình) Đây là một thuật ngữ chỉ một chương trình được thực thi và được sử dụng thay thế cho thuật ngữ tác vụ ("task"). "Process" là từ được sử dụng trong các hệ điều hành như UNIX. Nó được sử dụng thường xuyên trong những năm gần đây

6. Tất cả các quá trình được hoàn thành → Tác vụ được kết thúc.

**Dispatching** là một bước chọn lựa một tác vụ có quyền ưu tiên cao nhất trong số các tác vụ nằm trong trạng thái sẵn sàng và chuyển nó tới trạng thái **running**.[55]
**Supervisor call (lời gọi hệ thống)** là một yêu cầu một chức năng của hệ điều hành, trong biểu đồ dịch chuyển trạng thái tác vụ, nó được biết như là một lệnh vào/ra. Ngắt vào/ra thông báo cho biết thao tác vào/ra đã hoàn thành.[56]

# Ngắt

Ngắt là việc treo tạm thời một chương trình hiện thời đang được thực thi vì bất cứ lí do gì và chuyển quyền điều khiển cho hệ điều hành để thực thi chương trình xử lý cần thiết. Có các ngắt ngoài gây bởi các trạng thái đặc trưng của phần cứng và các ngắt trong cố tình khi chương trình điều khiển được gọi từ trong một chương trình.

Khi CPU phát hiện ra một ngắt, hệ điều hành tiếp nhận ngắt đó, thay đổi chương trình đang thực hiện đến trạng thái cần thiết ưu tiên ngắt, kiểm tra nguyên nhân ngắt, và chuyển điều khiển đến chương trình con xử lí ngắt tương ứng. Vùng nhớ mà trạng thái này được lưu trữ được gọi là từ trạng thái chương trình PSW (program status word). Cũng có thể xảy ra trường hợp trong khi một xử lý ngắt đang diễn ra, một ngắt khác lại xảy ra. Khi đó các ngắt được điều khiển dựa trên quyền ưu tiên của chúng. Bảng sau mô tả các loại ngắt chính và ví dụ nguyên nhân sinh ngắt.

| Loại ngắt | Tên | Các nguyên nhân có thể |
|---|---|---|
| Các ngắt ngoài | Ngắt kiểm tra phần cứng | Lỗi chức năng của các khối, vấn đề về nguồn/điện áp. |
| | Ngắt định thời (timer) | Khoảng thời gian định thời để đạt được một lượng thời gian đã định ra. |
| | Ngắt vào/ra[57] | Hoàn thành việc vào/ra. Khối vào/ra thay đổi trạng thái. (v.v…) |
| | Ngắt tín hiệu ngoài | Các lệnh từ bảng điều khiển, các tín hiệu ngoài. |
| Các ngắt trong (traps)[58] | Ngắt chương trình | Tràn trên, tràn dưới, thực thi các mã lệnh không được định nghĩa, chia cho 0, vi phạm vùng nhớ không cho phép. |
| | Lời gọi hệ thống | Các lệnh thao tác vào/ra, chuyển đổi thực hiện tác vụ, lỗi trang, lời gọi chương trình điều khiển. |

[55] **Dispatcher**: Là chương trình của hệ điều hành thực hiện dispatching (đẩy tác vụ từ hàng đợi vào trạng thái chạy), cũng được biết như là một chương trình con dispatching.
[56] (FAQ) Biểu đồ dịch chuyển trạng thái của các tác vụ hầu như luôn xuất hiện trong các bài thi. Cần nhớ kỹ biểu đồ này trong trí nhớ.
[57] (Gợi ý) Việc phát một lệnh vào/ra được thông báo bằng lời gọi hệ thống; Việc hoàn thành vào/ra được thông báo bởi một ngắt vào/ra.
[58] Các ngắt trong thường là cố tình bởi các chương trình, vì thế thỉnh thoảng chúng được gọi là các bẫy ngắt (traps)

# 2.2.4 Quản lí dữ liệu và tổ chức tệp

| Điểm chính | ➢ Quản lí dữ liệu cung cấp các phương pháp kết hợp để truy cập các tệp.<br>➢ Tổ chức tệp bao gồm tổ chức tuần tự, tổ chức trực tiếp, tổ chức theo chỉ số, tổ chức phân vùng, v.v… |
|---|---|

Một chức năng khác của hệ điều hành là "quản lí dữ liệu". Quản lí dữ liệu là chương trình điều khiển quản lí các dữ liệu vào và ra. Nó cung cấp các phương pháp tổ chức tệp khác nhau như tổ chức tuần tự, tổ chức trực tiếp, tổ chức chỉ số. Nó làm việc như một cầu kết nối giữa các tệp logic được xử lý trong một chương trình và các tệp vật lí có cấu trúc khác nhau.

Quản lí dữ liệu cho phép người lập trình không phải quan tâm đến cấu trúc vật lí của các tệp.

## Các phương pháp truy cập

Các phương pháp truy cập bao gồm truy cập tuần tự, truy cập trực tiếp, truy cập động (kết hợp cả 2 phương pháp trên).

| | |
|---|---|
| Truy cập tuần tự: | Xử lí các tệp tuần tự từ điểm bắt đầu[59] |
| Truy cập trực tiếp: | Xử lí trực tiếp một bản ghi cụ thể |
| Truy cập động: | Sử dụng truy cập trực tiếp để tìm và định vị một bản ghi, sau đó sử dụng truy cập tuần tự. |

Các

### y cập tuần tự

Đây là phương pháp truy cập các bản ghi trong một tệp theo trật tự tuần tự từ điểm bắt đầu. Có thể được thực hiện với hầu hết tất cả các phương tiện lưu trữ. Phương pháp này phù hợp cho xử lý chung tất cả các bản ghi trong một tệp cùng chủ đề để xử lý.

### Truy cập trực tiếp

Đây là phương pháp truy cập trong đó một bản ghi cần thiết được truy cập trực tiếp (một cách ngẫu nhiên) không quan tâm đến thứ tự bản ghi được lưu trữ. Phương pháp này được sử dụng khi phương tiện tệp là một phương tiện lưu trữ có thể truy cập trực tiếp ví dụ như đĩa cứng. Nó được sử dụng trong các hệ thống trực tuyến thời gian thực nơi chỉ có một phần của một

[59] (Gợi ý) Truy cập tuần tự có thể được sử dụng với hầu hết các phương tiện lưu trữ; còn truy cập trực tiếp và truy cập động bị hạn chế với các thiết bị có khả năng truy cập trực tiếp như đĩa cứng.

số lượng lớn các bản ghi được lưu trữ trong một tệp cần được truy cập để cập nhật nhanh chóng.

### Truy cập động

Đây là phương pháp truy cập trong đó sử dụng truy cập trực tiếp để tìm một bản ghi cụ thể và sau đó sử dụng truy cập tuần tự. Tương tự như truy cập trực tiếp, phương pháp này được sử dụng khi phương tiện lưu trữ tệp là một phương tiện lưu trữ có khả năng truy cập trực tiếp.

## Tổ chức tệp

Các phương pháp tổ chức tệp bao gồm: tổ chức tuần tự, tổ chức trực tiếp, tổ chức theo chỉ số và tổ chức phân vùng[60], v.v…

| | |
|---|---|
| Các tệp tổ chức tuần tự | Các tệp được xử lý theo thứ tự như chúng được ghi trong phương tiện lưu trữ. |
| Các tệp tổ chức trực tiếp | Các bản ghi được xử lý trực tiếp qua các giá trị khóa. |
| Các tệp tổ chức theo chỉ số | Truy cập trực tiếp qua chỉ số và có thể kèm truy cập tuần tự. |
| Các tệp tổ chức phân vùng | Một tệp chứa nhiều tệp tổ chức tuần tự. |

### C tệp tổ chức tuần tự

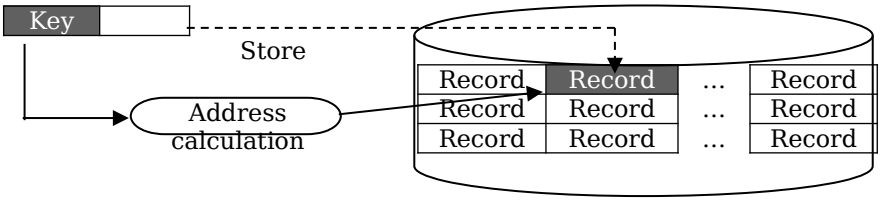
C bản ghi trong một tệp được lưu trữ tại những vị trí liên tiếp nhau theo m trật tự nào đó. Các tệp loại này có thể được thấy ở hầu hết các loại t bị lưu trữ như băng từ, đĩa cứng, đĩa mềm. Nhìn chung, chỉ có phương pháp truy cập tuần tự là có thể với các tệp loại này.

### Các tệp tổ chức trực tiếp

Địa chỉ lưu trữ trên phương tiện lưu trữ được tính toán dựa trên giá trị khóa để tìm ra địa chỉ mỗi bản ghi và bản ghi được lưu tại vị trí đó.[61] Để truy cập một bản ghi, đầu tiên phải tính địa chỉ lưu trữ theo cùng một công

[60] (FAQ) Các câu hỏi liên quan đến đặc trưng của các phương pháp tổ chức tệp thường được đưa ra. Vì vậy, cần nắm được đặc trưng của mỗi phương pháp tổ chức tệp.
[61] (Gợi ý) Một trường hợp đặc biệt của tổ chức trực tiếp là tổ chức tương đối, trong đó các giá trị khóa là liên tục ví dụ như 1, 2, 3, … và giá trị khóa cũng là địa chỉ lưu trữ của bản ghi.

thức và bản ghi được đọc tại địa chỉ lưu trữ đó. Giá trị khóa của mỗi bản ghi cũng có thể được sử dụng trực tiếp như là địa chỉ lưu trữ của bản ghi đó, nhưng điều này không thiết thực lắm vì nó tạo ra nhiều vùng nhớ rác nếu các giá trị khóa là không liên tục. Các thức nói chung là sử dụng một loại công thức biến đổi để tính ra địa chỉ lưu trữ từ giá trị khóa của mỗi bản ghi. Cách này được gọi là phép biến đổi địa chỉ (một cách ngẫu nhiên).




Phép biến đổi địa chỉ đôi khi sinh ra cùng một địa chỉ cho những bản ghi khác nhau. Trong trường hợp như vậy, bản ghi được lưu trữ đầu tiên được gọi là bản ghi chính trong khi bản ghi được đăng kí sau tại cùng địa chỉ đó được gọi là bản ghi đồng nghĩa.[62]

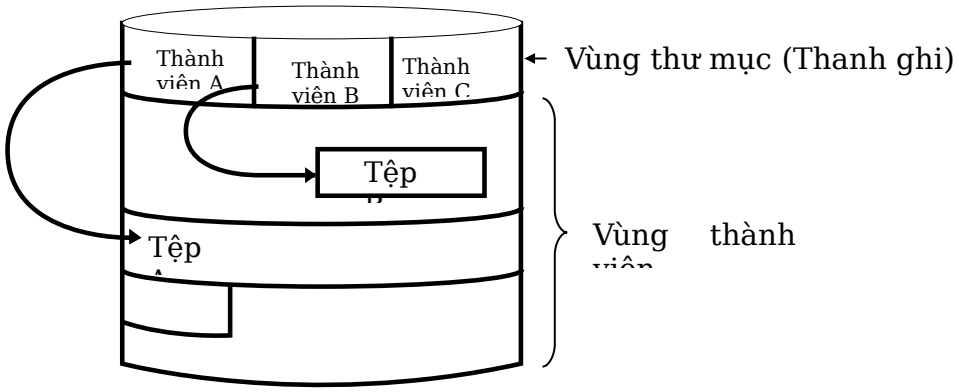
[62] **Synonym/Home:** Phép biến đổi địa chỉ có thể sinh ra các giá trị khóa khác nhau nhưng cùng một giá trị lưu trữ như nhau, gọi là một "đồng nghĩa". Nếu kết quả của phép biến đổi khóa lưu trữ một bản ghi, bản ghi này được gọi là bản ghi chính, và bản ghi mà không thể được lưu trữ gọi là bản ghi đồng nghĩa. Một bản ghi đồng nghĩa cần được lưu trữ tại bất cứ nơi nào bằng một phương pháp khác (ví dụ dùng một danh sách).

### Các tệp tổ chức theo chỉ số

Mỗi tệp sẽ có một chỉ số, và chúng được tổ chức để người sử dụng có thể truy cập đến các bản ghi bằng cách tìm địa chỉ của các bản ghi đó qua chỉ số tệp. Các phương pháp truy cập tuần tự, truy cập trực tiếp, và truy cập động đều có thể được dùng. Trong mỗi trường hợp, các bản ghi thực tế được truy cập chỉ sau khi đã tìm được địa chỉ của chúng trên phương tiện lưu trữ thông qua chỉ số, vì thế phương tiện lưu trữ không chỉ chứa một vùng dữ liệu cơ sở nơi lưu trữ dữ liệu mà nó còn chứa một vùng chỉ số. Thêm nữa, để ngăn chặn tình huống các bản ghi không thể ghi thêm vào vùng dữ liệu cơ sở, có thêm một vùng nhớ tràn để dự trù. [63]

### Các tệp tổ chức phân vùng

Trong các tệp này, các tệp tổ chức tuần tự được gộp lại vào các đơn vị gọi là các thành viên, mỗi một thành viên được cung cấp một tên. Sau đó một thư mục chứa các tên đó và địa chỉ đầu của chúng được tạo ra. Việc truy cập được cho phép đến các thành viên đó. Mỗi thành viên được xem như một tập các tệp tổ chức tuần tự. Phương pháp truy cập trực tiếp có thể tìm thấy chỗ bắt đầu của một thành viên, và phương pháp truy cập tuần tự có thể được sử dụng để tìm một bản ghi trong thành viên đó. Các tệp tổ chức phân vùng có thể được sử dụng cho các tệp chương trình và thư viện nhưng không được dùng nhiều như các tệp dữ liệu.



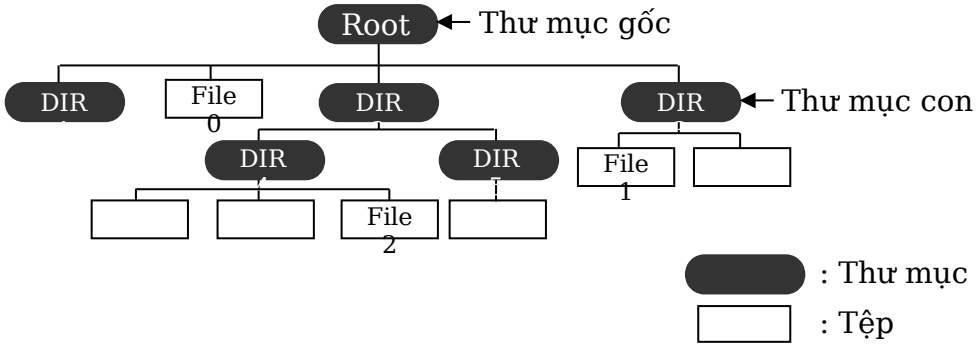

[63] (Gợi ý) Có 2 loại vùng tràn: vùng tràn cylinder và vùng tràn độc lập. Các bản ghi tràn từ các track khác nhau được lưu trữ trong vùng tràn cylinder; Nếu vùng tràn cylinder đầy, các bản ghi sẽ được lưu trữ thêm vào một vùng tràn độc lập được dùng chung cho tất cả các tệp.

# Hệ thống tệp phân cấp

UNIX và Windows[64] sử dụng các hệ thống tệp phân cấp như một cơ chế để quản lí tệp một cách hiệu quả.

### Thư mục và tệp

Một hệ thống tệp có cấu trúc phân cấp bao gồm các tệp và các thư mục (các thư mục là các thanh ghi tệp). Đỉnh của cấu trúc phân cấp là thư mục gốc, và các thư mục dưới nó được gọi là các thư mục con[65]




### Các thao tác tệp

Khi tìm kiếm một tệp, cần chỉ ra đường dẫn thư mục nơi mà tệp đó được lưu trữ. Có 2 phương pháp để thực hiện điều này. Ví dụ với trường hợp cấu trúc phân cấp như hình trên, ta có các đường dẫn như dưới đây.[66]

- Đường dẫn tuyệt đối

  Chỉ rõ một đường dẫn từ thư mục gốc.[67]
  <Ví dụ> Đây là đường dẫn đến "file1"
  \DIR3\file1 (Ký tự đầu tiên "\" chỉ định thư mục gốc.)

- Đường dẫn tương đối

  Chỉ rõ một đường dẫn từ thư mục hiện thời[68]
  <Ví dụ 1> Đây là đường dẫn của file2 khi thư mục hiện thời là "DIR2."
  DIR4\file2
  <Ví dụ 2> Đây là đường dẫn của file2 khi thư mục hiện thời là

[64] (Gợi ý) Thuật ngữ "directory" trong UNIX và MS-DOS tương ứng với "folder" trong Windows và MacOS.
[65] (Gợi ý) Các tệp và thư mục có thể được tạo trong một thư mục, nhưng không thể trong một tệp.
[66] (FAQ) Liên quan đến các hệ thống tệp phân cấp, có các câu hỏi thi giống như "Chọn một đường dẫn tuyệt đối hoặc đường dẫn tương đối". Trong các câu hỏi này, ký tự phân tách các thư mục và các tệp được sử dụng sẽ được giải thích trong câu hỏi.
[67] (Gợi ý) Ở đây chúng ta sử dụng ký tự "\" để phân tách các thư mục và các tệp, tuy nhiên một vài hệ điều hành sử dụng ký tự "/".
[68] **Thư mục hiện thời** (**Current directory)**: Thư mục hiện thời là thư mục trong đó người sử dụng hiện tại đang làm việc.

"DIR4."

file2

## 2.2.5 Quản lý bộ nhớ

| Điểm chính | ➢ Quản lý bộ nhớ sử dụng 2 loại bộ nhớ: bộ nhớ thực và bộ nhớ ảo.<br>➢ Định dạng cơ sở của bộ nhớ ảo là phương pháp phân trang. |
|---|---|

Một vai trò khác của hệ điều hành là "quản lý bộ nhớ". Quản lý bộ nhớ đem lại hiệu quả sử dụng bộ nhớ cao nhất cũng như hạn chế việc thiếu hụt dung lượng bộ nhớ. Các bộ nhớ phụ cũng được sử dụng hiệu quả như một phần của bộ nhớ chính.[69]

### Hệ thống bộ nhớ thực[70]

Đây là một hệ thống mà quản lý không gian vật lý của bộ nhớ chính. Bộ nhớ thực có thể được điều khiển bằng nhiều cách: phân vùng, hoán đổi, bố trí lại và các phương pháp ghi đè.

#### Phương pháp phân vùng

Khi một chương trình được đặt trong bộ nhớ chính, bộ nhớ chính được phân chia thành một số phân vùng nơi mà chương trình được nạp.[71] Nếu không có quản lý bộ nhớ, sự phân mảnh sẽ tạo ra gây nên tình trạng cản trở các chương trình được lưu trữ mặc dù không gian trống vẫn còn. Do đó, việc kết hợp các vùng nhớ trống với nhau là cần thiết.[72]

#### Hoán đổi (roll-in / roll-out)

Hoán đổi là việc chương trình thực hiện chuyển đổi thứ tự sau và trước giữa bộ nhớ chính và bộ nhớ phụ. Nếu một chương trình được đưa vào có quyền ưu tiên cao hơn quyền ưu tiên của chương trình thực thi hiện thời, chương trình mới sẽ ngay lập tức được nạp vào bộ nhớ chính và được thực thi. Tuy nhiên, nếu không gian nhớ trống trong bộ nhớ chính không còn, một chương trình bất kỳ trong bộ nhớ chính có thể được dịch chuyển ra ngoài bộ nhớ phụ. Do đó, hệ thống sẽ bù đắp được sự thiếu hụt dung lượng bộ nhớ chính bằng cách sử dụng bộ nhớ phụ ngoài. Tuy nhiên, nếu hoán đổi xảy ra thường xuyên, có nghĩa là chương trình đó được chuyển đổi thứ tự trước sau nhiều lần, như thế sẽ làm giảm thiểu hiệu quả xử lý của hệ thống máy tính.

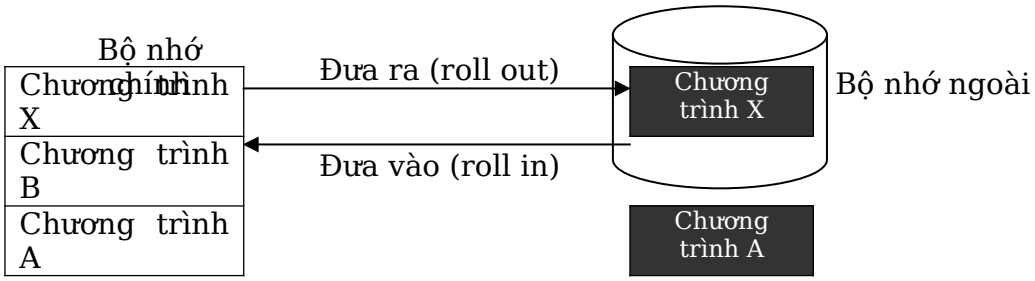
---

[69] **Rò rỉ bộ nhớ** (**Memory leak)**: Đôi khi vì một số lý do, vùng nhớ trong bộ nhớ chính tự động bị chiếm dụng bởi một chương trình ứng dụng, và không được giải phóng mà vẫn còn trong bộ nhớ chính. Đây là hiện tượng rò rỉ bộ nhớ (Memory leak). Để loại trừ hiện tượng này, cần thực hiện việc nén bộ nhớ.

[70] **Hệ thống nhớ thực hay Bộ nhớ thực** (**Real memory system or "Real Storage (RS)")**: Đây là bộ nhớ tồn tại thực sự; nó là bộ nhớ chính.

[71] (Gợi ý) Trong phương pháp phân vùng, nhiều chương trình có thể được lưu trữ đồng thời, vì thế có thể xử lý đa nhiệm.

[72] **Nén bộ nhớ (Compaction)**: Nghĩa là thu gom các vùng nhớ trống để được một vùng nhớ liên tục; cũng được gọi là thu gom bộ nhớ rác.


Bộ nhớ chính
Chương trình X
Chương trình B
Chương trình A
Đưa ra (roll out)
Đưa vào (roll in)
Chương trình X
Chương trình A
Bộ nhớ ngoài

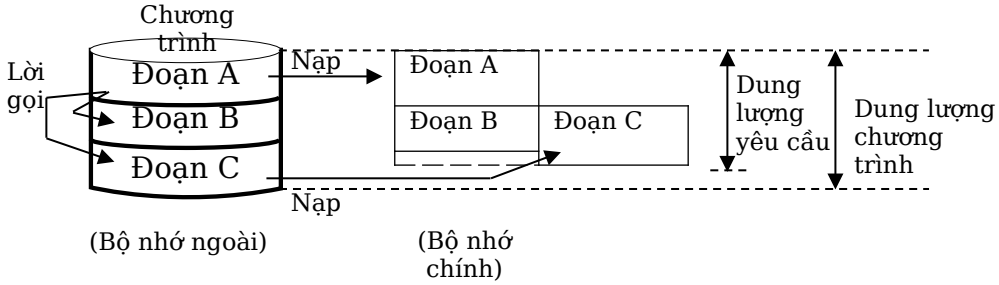
### Bố trí lại (Relocation)

Bố trí lại là chức năng trong đó một chương trình đã được đăng kí một vùng nhớ được lưu trữ lại ở một vùng nhớ khác. Một chương trình mà cấu trúc của nó cho phép thực hiện bố trí lại được gọi là chương trình có khả năng bố trí lại.[73]

### Phương pháp ghi đè

Giới hạn vật lý của bộ nhớ chính có thể được loại bỏ, nhờ vào việc các chương trình được phân thành các đơn vị đoạn, và chỉ có các đoạn[74] cần thiết được nạp vào bộ nhớ chính để thực thi. Toàn bộ chương trình được lưu trữ ở bộ nhớ ngoài, bộ nhớ chính chỉ chứa các đoạn thường sử dụng. Các đoạn khác mà không bao giờ được sử dụng đồng thời, được nạp từ bộ nhớ ngoài vào bộ nhớ chính khi được yêu cầu đến.

Ví dụ, giả sử rằng đoạn A là một thủ tục chính được sử dụng với tần số cao trong khi các đoạn B và C là các thủ tục con được gọi riêng biệt từ đoạn A. Trong khi đoạn B được thực thi, đoạn C vẫn lưu trữ trong bộ nhớ ngoài. Khi đoạn C được gọi, nó sẽ được nạp vào vùng nhớ của đoạn B. Do đó, toàn bộ dung lượng bộ nhớ chính là "A + B + C", nhưng thực tế dung lượng của bộ nhớ chính có đủ chỉ là giá trị lớn nhất của "A+B" hoặc "A+C".




# Bộ nhớ ảo

Bộ nhớ ảo cung cấp một dung lượng lớn không gian lưu trữ mà không quan tâm đến kích thước của bộ nhớ chính.[75] Các chương trình được lưu trữ trong bộ nhớ ảo (thông thường trong bộ nhớ ngoài), và chỉ các phần cần thiết để thực thi mới được nạp vào bộ nhớ chính.

Khi chương trình được nạp vào bộ nhớ ảo, các lệnh và dữ liệu được cấp các địa chỉ ảo, cái mà cần thiết phải biến đổi sang các địa chỉ thực (các địa chỉ trong bộ nhớ chính) cho việc thực thi của chương trình. Phép biến đổi này được thực hiện bằng phần cứng gọi là bộ dịch địa chỉ động DAT (Dynamic Address Translator).

Nếu bộ nhớ ảo được sử dụng, cần phải biến đổi từ các địa chỉ bộ nhớ ảo sang

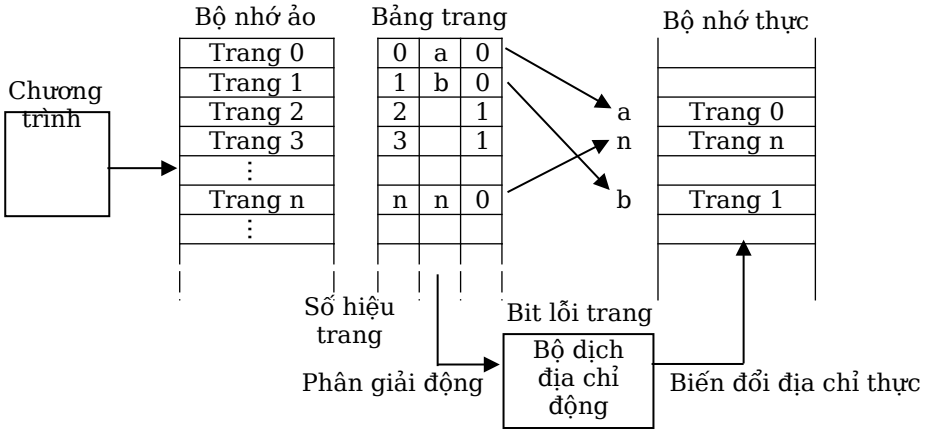
[73] (Gợi ý) Thuật ngữ "Relocatable" (có thể bố trí lại) có nghĩa là có thể nén (gom lại).
[74] **Segment**: (Đoạn) là một đơn vị xử lý logic của chương trình. Ở đây, chúng ta có thể hiểu một đoạn như một chương trình con.
[75] **Virtual memory system or "Virtual Storage (VS)"**: Đây là một khái niệm lưu trữ không tồn tại thực sự. Một chương trình cần thực thi được nạp vào bộ nhớ ảo, có không gian bộ nhớ lớn hơn, trong đó chỉ các phần (các trang và các đoạn) của chương trình với tần số sử dụng cao, dữ liệu và các phần cần thiết khác để thực thi được nạp vào bộ nhớ chính.

địa chỉ vật lý trong bộ nhớ chính. Phép biến đổi địa chỉ này được thực hiện ở tốc độ cao bởi DAT. Sự bố trí lại cũng được thực hiện nếu bộ nhớ chính có sự phân đoạn. Khi sự bố trí lại hoàn thành trong suốt quá trình thực hiện của một chương trình, nó được gọi là sự bố trí lại động.

Bộ nhớ ảo sử dụng 3 phương pháp: phân trang, phân đoạn và phương pháp kết hợp.

### Phương pháp phân trang

Trong phương pháp này, chương trình được phân thành các khối có kích thước cố định gọi là các trang. Mỗi trang sẽ trở thành một khối để nạp vào bộ nhớ thực. Các trang được quản lý bởi một bảng phân trang, trong đó mỗi một mục tương ứng với một trang trong bộ nhớ ảo. Nếu trang tương ứng nằm trong bộ nhớ thực, bit lỗi trang sẽ bằng 0. Bit lỗi trang để chỉ ra có hay không trang tương ứng trong bộ nhớ thực.




### Phương pháp phân đoạn

Trong phương pháp này, các chương trình và các tập hợp logic của dữ liệu được xem như là các đoạn. Các địa chỉ ảo bao gồm các số hiệu đoạn và các địa chỉ trong các đoạn đó. Phương pháp phân trang chỉ dùng cho quản lý bộ nhớ và như vậy các chương trình không cần ghi nhớ các trang đó. Ngược lại, trong phương pháp phân đoạn, các đoạn có kích thước khác nhau, các chương trình phải ghi nhớ kích thước của các đoạn.

Các đoạn là các đơn vị xử lý logic, vì thế chúng có thể được đối xử như là các thủ tục con. Tuy nhiên, việc chiều dài thay đổi thường gây khó khăn cho việc quản lý và hiệu năng sử dụng bộ nhớ chính có thể bị suy giảm.

### Phương pháp kết hợp phân trang – phân đoạn

Đây là một cải tiến của phương pháp phân đoạn, trong đó các đoạn được phân nhỏ hơn thành các trang. Các địa chỉ thật được truy cập theo trật tự "đoạn → trang → độ dịch chuyển tương đối trong trang".[76]

[76] **Relative displacement within the page**: (Độ dịch chuyển tương đối trong trang) Là một địa chỉ được đăng kí để bắt đầu trang có độ dịch chuyển là 0.

# Các giải thuật phân trang

Nếu một trang cần xử lý không tìm thấy trong bộ nhớ thực, một ngắt gọi lỗi phân trang xảy ra, và trang đó được đọc từ bộ nhớ thực vào bộ nhớ ảo. Quá trình này được gọi là nạp trang (page-in). Trái lại, loại trang (page-out) là dịch chuyển một trang không cần thiết ra khỏi bộ nhớ ảo. Các thao tác nạp trang và loại trang được gọi là phân trang (paging).[77]

Nếu phân trang xảy ra thường xuyên, thời gian để thực thi chương trình điều khiển tăng, giảm hiệu quả thực hiện. Điều này gọi là hiện tượng slashing. Để giảm thiểu khả năng xảy ra hiện tượng slashing nhiều nhất có thể, nhiều giải thuật khác nhau được đề xuất để chọn trang.

Các phương pháp loại trang (page-out) phổ biến và đặc điểm của chúng được chỉ ra như dưới đây.[78]

| Phương pháp | Đặc điểm |
|---|---|
| LRU (Least Recently Used) | Các trang được so sánh thời gian tính từ lần tham chiếu cuối cùng. Trang nào có thời gian này nhiều nhất sẽ được lấy ra |
| FIFO (First In First Out) | Trang nào có thời gian tính từ lúc nạp vào đến thời điểm hiện tại dài nhất sẽ được lấy ra trước (vào trước ra trước). |

[77] (Gợi ý) Hoán đổi (Swapping) và phân trang (paging) là giống nhau, nhưng chú ý rằng hoán đổi diễn ra trong chương trình và các đoạn trong khi phân trang diễn ra trong các khối gọi là các trang, là một phần của chương trình.

[78] (FAQ) Có các câu hỏi thi yêu cầu phân định rõ ràng nạp trang, loại trang. Ví dụ, nếu thứ tự các trang được sử dụng là “1, 3, 2, 3, 5, 2” và nếu bộ nhớ chính có trang 3, trang nào sẽ sẽ được loại trước ? Với các câu hỏi loại này, cần hiểu rõ các giải thuật LRU và FIFO.

## Câu hỏi nhanh

**Q1** Giải thích các vai trò của quản lí tác vụ

**Q2** Kiểu tổ chức sau đây thuộc kiểu tổ chức tệp nào ? Các tệp tổ chức tuần tự được phân nhóm thành các khối gọi là các thành viên, mỗi thành viên được cấp một tên. Một thư mục được tạo ra bao gồm các tên và các địa chỉ đầu và cho phép truy cập đến các thành viên đó.

**Q3** Giải thích kỹ thuật hoán đổi (swapping)

**Q4** Đơn vị gì được nạp vào bộ nhớ chính trong phương pháp phân trang ?

**A1** Đây là một chức năng điều khiển việc thực thi chương trình, thực hiện các chức năng như điều khiển đồng bộ các chương trình, cấp phát động các tài nguyên để thực hiện các chương trình, và quản lý quyền ưu tiên của các chương trình. Nó cũng có thể kiểm soát rất nhiều các điều khiển ngắt.

**A2** Tổ chức phân vùng

**A3** Hoán đổi (swapping) là việc chương trình thực hiện chuyển đổi thứ tự trước và sau giữa bộ nhớ chính và bộ nhớ ngoài. Nếu một chương trình được đưa vào với quyền ưu tiên cao hơn mức ưu tiên của chương trình thực thi hiện thời, chương trình mới sẽ ngay lập tức được nạp vào bộ nhớ chính, bất cứ chương trình nào trong bộ nhớ chính cũng có thể được di chuyển ra bộ nhớ ngoài. Do đó, hệ thống sẽ bù đắp được sự thiếu hụt không gian nhớ bằng cách sử dụng bộ nhớ ngoài.

**A4** Trang (page).

# 2.3 Kỹ thuật cấu hình hệ thống

## Giới thiệu

Nhiều cấu hình hệ thống khác nhau được sử dụng để giảm thiểu chi phí và tăng hiệu năng của hệ thống máy tính. Bao gồm các hệ thống khách/chủ để phân tán gánh nặng, các hệ thống song hành để tăng độ tin cậy, và các hệ thống kép (dự phòng).

## 2.3.1 Các hệ thống khách chủ

| Điểm chính | ➢ Hệ thống khách/chủ là một ví dụ thông thường về xử lý phân tán.<br>➢ Các máy khách yêu cầu xử lý và các máy chủ cung cấp các dịch vụ (xử lý). |
|---|---|

Hệ thống khách chủ (Client Server System (CSS)) là một hệ thống máy tính trong đó sử dụng một mạng để xử lý phân tán; Đây là một kiểu cấu hình hệ thống.[79]

### Tổng quan về hệ thống khách chủ

CSS được cấu hình với các máy tính đảm nhận các vai trò gọi là máy khách và máy chủ (máy phục vụ). Các máy chủ cung cấp các dịch vụ (xử lý) như quản lý tệp, quản lý cơ sở dữ liệu, sửa đổi và cung cấp dữ liệu, điều khiển in ấn, và các chức năng giao tiếp, được yêu cầu bởi các máy khách.

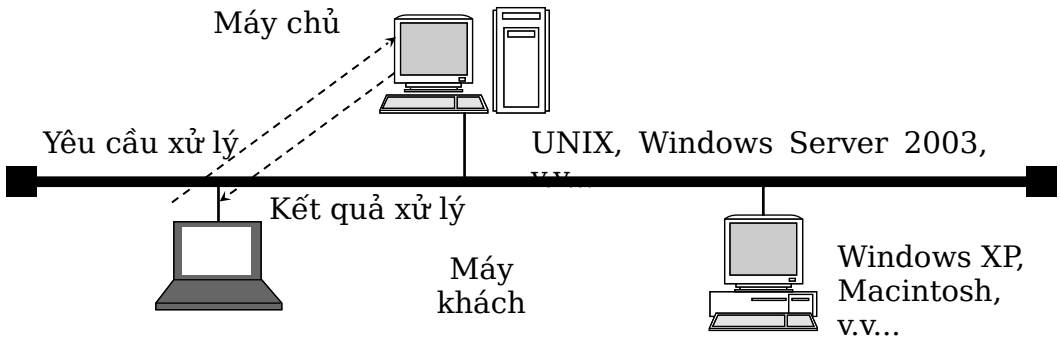
Máy khách gửi yêu cầu dịch vụ đến máy chủ, nhận các kết quả dữ liệu được xử lý bởi máy chủ, và hiển thị các kết quả đó.

Các máy khách và các máy chủ phân tán các xử lý của chúng để cố gắng làm giảm gánh nặng trong việc xử lý của các máy. Ngoài ra, bằng cách chia sẻ các tài nguyên, người sử dụng có thể giảm thiểu được chi phí thừa. Ví dụ, bằng cách kết nối một máy in tốc độ cao với máy chủ, các máy khách có thể cùng chia sẻ máy in tốc độ cao đó. Chi phí cho một máy in tốc độ cao có thể rẻ hơn việc chuẩn bị cho mỗi máy khách một máy in tốc độ thấp, phụ thuộc vào số lượng máy khách.

Các máy tính được sử dụng như một máy chủ thường có hiệu năng cao hơn các máy khách. Để phân loại các chức năng của các máy chủ khác nhau, có thể gọi tên theo chức năng của chúng, như máy chủ tệp, máy chủ cơ sở dữ liệu, máy chủ in, máy chủ giao tiếp.[80]

[79] Hệ thống khách/chủ: Trong lập trình đa xử lý, người sử dụng có thể chạy cả dịch vụ khách và chủ trong một máy tính. Mặt khác xử lý khách/chủ không có nghĩa là mọi thứ được phân tán. Đây là một phương pháp để đạt được xử lý phân tán.

[80] (Gợi ý) Nếu một yêu cầu dịch vụ của máy khách không được cung cấp bởi máy chủ phục vụ, máy chủ phục vụ có thể trở thành một máy khách và yêu cầu đến một máy chủ phục vụ khác để




# Các loại máy chủ

Bảng sau mô tả các loại máy chủ, dựa trên các chức năng cung cấp.[81]

| Các loại máy chủ | Các dịch vụ cung cấp |
|---|---|
| Máy chủ tệp | Quản lý các tệp được chia sẻ, điều khiển việc truy cập và ghi vào các tệp. |
| Máy chủ cơ sở dữ liệu | Quản lý cơ sở dữ liệu, thao tác cơ sở dữ liệu với DBMS (hệ thống quản lý cơ sở dữ liệu) |
| Máy chủ in | Quản lý máy in được chia sẻ, in ấn khi nhận được một yêu cầu in. |
| Máy chủ giao tiếp | Cung cấp các chức năng giao tiếp với bên ngoài sử dụng kết nối mạng. |

# Các đặc trưng của hệ thống khách/chủ

Bởi vì hệ thống khách/chủ cung cấp kiểu xử lý phân tán thông thường, vì vậy các đặc trưng của xử lý phân tán được áp dụng là điều dễ hiểu.[82]

### Các ưu điểm của hệ thống khách/chủ

- Khi các xử lý diễn ra tại các máy khách, việc phản hồi sẽ nhanh hơn.
- Các chi phí cho toàn bộ hệ thống thông tin có thể được giảm thiểu, đạt được tỷ lệ chi phí/hiệu năng tốt.
- Các chức năng máy chủ chuyên dụng cung cấp cho hệ thống hiệu năng và hiệu quả kinh tế cao hơn.
- Dễ dang mở rộng, các máy khách và máy chủ có thể được thêm vào một cách linh hoạt khi cần thiết.
- Thậm chí khi các máy khách truy cập đến một máy chủ, chúng cũng không biết là đang ở trong một môi trường phân tán.
-

### Các nhược điểm của hệ thống khách/chủ

- Hệ thống có thể gặp rắc rối nếu không có người quản trị máy chủ rõ ràng.

xử lý yêu cầu đó.

[81] (Gợi ý) Các máy khách và các máy chủ không cần phải chạy các hệ điều hành giống nhau. Ở những nơi có nhiều máy chủ, chúng không cần phải chạy các hệ điều hành giống nhau. Ngoài ra, nếu có nhiều máy khách, chúng cũng không cần chạy các hệ điều hành giống nhau.

[82] (Gợi ý) Nếu các chương trình khách/chủ được phân chia logic vào 3 tầng (tầng trình chiếu, tầng ứng dụng, và tầng dữ liệu) thì hệ thống như vậy được gọi là hệ thống khách/chủ 3 tầng. Bằng cách phân biệt 3 tầng theo chức năng như vậy, hệ thống cố gắng nâng cao hiệu năng và hiệu quả phát triển, bảo trì.

- Hiệu năng bị giảm thiểu nếu chỉ sử dụng tập trung trên các máy chủ chuyên dụng.
- Hiệu năng của toàn bộ hệ thống phụ thuộc vào hiệu năng của mạng.[83]

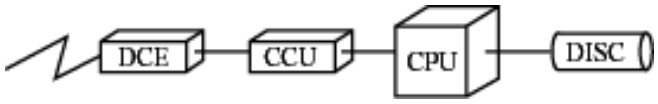
## 2.3.2 Các cấu hình hệ thống

| Điểm chính | ➢ Hệ thống song hành là một cấu hình đem lại sự tin cậy cao; hệ thống dự phòng là một cấu hình đem lại tính sẵn dùng cao.<br>➢ Có 2 loại hệ thống đa bộ xử lý: hệ thống cặp đôi lỏng lẻo và hệ thống cặp đôi chặt chẽ. |
|---|---|

Có nhiều mô hình cấu hình hệ thống có thể dùng căn cứ theo mục tiêu xử lý thông tin. Ví dụ, tính tin cậy của một hệ thống máy tính có thể được nâng cao bằng cách sử dụng nhiều khối. Chúng ta sẽ xem xét các cấu hình hệ thống chính và đặc trưng của chúng trên các phương diện tính tin cậy, hiệu quả, chi phí và các yếu tố khác.

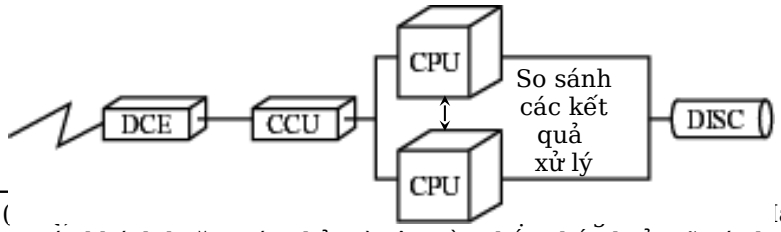
### Hệ thống đơn công

Hệ thống này chỉ bao gồm một CPU. Tính tin cậy và khả năng xử lý kém hơn so với các cấu hình khác, nhưng nó có tính kinh tế hơn. Đây là cấu hình được sử dụng phổ biến. [84]




### Hệ thống song hành

Đây là một cấu hình hệ thống mà có 2 CPU cùng thực hiện một xử lý và so sánh kết quả xử lý đó với nhau. Cấu hình này được ứng dụng khi việc xử lý đòi hỏi không bao giờ được phép dừng, thậm chí chỉ một lúc. Nếu một CPU bị lỗi, hệ thống ngắt CPU bị lỗi đó và tiếp tục xử lý trên CPU còn lại. Tính tin cậy là cực kỳ cao, nhưng điều đó cũng làm cho hệ thống trở nên rất đắt.[85] [86]




[83] ( ầu hết trong số đó là về vai trò của máy khách hoặc máy chủ. Vì vậy cần chắc chắn hiểu rõ các kiến thức về chúng.

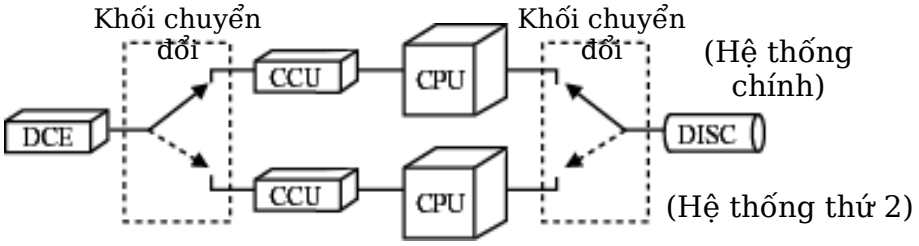
[84] **DCE (Data Circuit-terminating Equipment)**: Là thiết bị biến đổi các tín hiệu nhận được trên đường truyền thông , gửi chúng đến các thiết bị đầu cuối, và cũng thực hiện các thao tác ngược lại. Thông thường, thiết bị này được kết nối tại đầu cuối của đường truyền thông và có chức năng như một giao tiếp với máy tính. Một modem được sử dụng trên đường tín hiệu tương tự, và DSU (Digital Service Unit) được sử dụng trên đường tín hiệu số.

[85] **CCU (Communication control unit)**: Là khối điều khiển thu và phát dữ liệu, thực hiện điều khiển lỗi, và thu thập, phân tích các ký tự.

[86] **DISC**: Thiết bị lưu trữ ngoài

## Hệ thống kép (hệ thống dự phòng)

Hệ thống này có 2 CPU, trong đó hệ thống chính được sử dụng cho xử lý trực tuyến còn hệ thống phụ thứ 2 được sử dụng cho các xử lý có quyền ưu tiên thấp hơn như xử lý lô.[87] Nếu hệ thống chính bị lỗi, việc xử lý trực tuyến sẽ chuyển sang hệ thống thứ 2.[88] Thời gian chuyển từ vài chục phút đến vài giờ. Tính tin cậy của nó thấp hơn hệ thống song hành, nhưng hệ thống này có điểm tốt hơn ở chi phí, vì thế nó cũng thường được sử dụng hơn hệ thống song hành.




## Hệ thống đa bộ xử lý (Các bộ xử lý đồng thời)

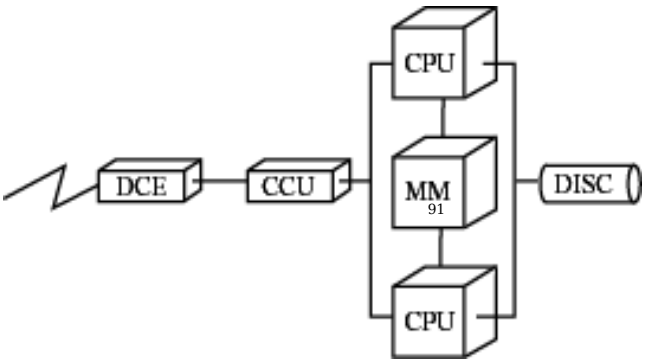
Trong hệ thống này, nhiều bộ CPU và CCU cùng chia sẻ và xử lý các tác vụ, vì thế đây là một cấu hình hệ thống có hiệu năng xử lý cao. Có 2 loại hệ thống đa bộ xử lý. Loại thứ nhất là LCMP[89] (loosely coupled multiprocessors – hệ thống cặp đôi lỏng lẻo), trong đó nhiều hệ thống máy tính được điều khiển bởi các hệ điều hành riêng biệt. Loại thứ 2 là TCMP[90] (tightly coupled multiprocessors – hệ thống cặp đôi chặt chẽ), trong đó nhiều hệ thống máy tính chia sẻ bộ nhớ chính và được điều khiển bởi cùng một hệ điều hành. Hình sau chỉ ra một ví dụ về TCMP.

[87] (FAQ) Có nhiều câu hỏi thi về đặc trưng của các hệ thống song hành, dự phòng và đa bộ xử lý. Thuật ngữ chính của mỗi hệ thống là như sau: “so sánh các kết quả xử lý” với hệ thống song hành, “Chuyển đổi giữa các khối” với hệ thống dự phòng, và “chia sẻ bộ nhớ chính” với hệ thống đa bộ xử lý.

[88] (Note) Nếu có một lỗi xảy ra trong hệ thống chính, sẽ mất thời gian để chuyển sang hệ thống thứ 2. Đó là bởi vì xử lý theo lô hoặc bất cứ việc gì đang được thực trong hệ thống thứ 2 phải được tạm dừng, và hệ điều hành phải được khởi động cho hệ thống trực tuyến. Một cấu hình hệ thống dự phòng nóng có thể giải quyết vấn đề này, sẵn sàng thay thế bất cứ khi nào. Trong trường hợp này, hệ điều hành trên hệ thống trực tuyến lưu lại, do đó việc chuyển đổi có thể diễn ra ngay lập tức.

[89] **LCMP (Loosely Coupled Multiprocessor)**: Mỗi một CPU có bộ nhớ chính riêng và hệ điều hành độc lập. Các CPU cùng tham ra vào một mạng tốc độ cao hoặc đường dẫn chia sẻ. Đây là một cấu hình trong đó các hệ thống máy tính độc lập được kết nối qua một mạng.

[90] **TCMP (Tightly Coupled Multiprocessor)**: Một bộ nhớ chính và một hệ điều hành được chia sẻ chung trong cấu hình này. Mỗi CPU có thể thực hiện các xử lý giống nhau, vì thế nếu một CPU bị hỏng, xử lý vẫn có thể tiếp tục, mặc dù hiệu năng thấp hơn. Cấu hình này có tính tin cậy cao và vì thế nó được sử dụng trong các hệ thống yêu cầu khả năng xử lý ở mức độ cao.




[91] **MM**: Bộ nhớ chính

# 2.3.3Xử lý tập trung và xử lý phân tán

| Điểm chính | ➢ Xử lý tập trung có mức độ an toàn cao nhưng lại không linh hoạt.<br>➢ Xử lý phân tán có tính kinh tế, nhưng khó hiểu rõ về bản chất. |
|---|---|

Dựa vào việc các máy tính giữ vị trí như thế nào về mặt vật lý, có 2 kiểu xử lý: Xử lý tập trung và xử lý theo phân tán.

## Xử lý tập trung

Xử lý tập trung là một cấu hình hệ thống trong đó một máy tính được kết nối với nhiều thiết bị cuối, và máy tính đó một mình thực hiện tất cả các xử lý. Dễ dàng trong việc duy trì tính nhất quán dữ liệu và cũng dễ dàng để quản lý tài nguyên. Những ưu điểm đó làm cho cấu hình này trở nên phổ biến ở những nơi mà một máy tính mục đích chung được sử dụng như một máy chủ trong xử lý tập trung. Dưới đây tóm lược sự so sánh tương đối với xử lý phân tán.

| Ưu điểm của xử lý tập trung | Nhược điểm của xử lý tập trung |
|---|---|
| Dễ dàng cải thiện tỷ lệ chi phí/ hiệu năng. (Grosch's Law[92])<br>Vận hành và bảo trì đòi hỏi ít nhân viên hơn<br>Độ an toàn của hệ thống cao | Khả năng mở rộng kém để thích hợp với các công nghệ mới.<br>Dễ xảy ra sự tắc nghẽn.[93]<br>Chi phí cho hệ điều hành là rất quan trọng.<br>Khôi phục một lỗi máy chủ sẽ tiêu tốn thời gian<br>Một hỏng hóc có thể gây ảnh hưởng sâu rộng. |

## Xử lý phân tán

Xử lý phân tán là một cấu hình hệ thống trong đó nhiều máy tính kết nối với nhau qua một mạng để thực hiện việc xử lý. Bởi vì việc xử lý được thực hiện qua một mạng nên thời gian xử lý sẽ lâu hơn so với xử lý tập trung. Tuy nhiên, nó có ưu điểm là nếu xảy ra lỗi ở một máy thì vẫn không ảnh hưởng đến toàn bộ hệ thống. Dưới đây sẽ tóm lược sự so sánh tương đối với xử lý tập trung.[94]

| Các ưu điểm của xử lý phân tán | Các ưu điểm của xử lý tập trung |
|---|---|
| Các trách nhiệm quản lý là rõ ràng. (Các trách nhiệm quản lý có thể được phân cho mỗi tổ chức.)<br>Ảnh hưởng của lỗi hệ thống chỉ mang tính cục bộ. | Trạng thái thực tế khó nắm bắt.<br>Khó xác định vấn đề lỗi ngay lập tức.<br>Hiệu năng của mạng có ảnh hưởng lớn..<br>Dễ xảy ra sự không nhất quán dữ liệu. |

[92] **Grosch's Law**: Phát biểu rằng "Hiệu năng tỷ lệ thuận với bình phương của giá trị". Nếu giá của máy tính tăng gấp đôi, hiệu năng tăng gấp 4. Tuy nhiên, tiến bộ về công nghệ đã giảm thiểu đáng kể giá cả của các thiết bị, vì vậy định luật này không còn được áp dụng.

[93] **Backlog**: Nghĩa là các hệ thống, phần mềm, các chương trình, v.v... cần thiết để phát triển nhưng sự phát triển đó chưa được bắt đầu. Thuật ngữ này thường chỉ những thứ mà được tổ chức trong phòng công nghệ thông tin trong một công ty.

[94] (FAQ) Nhiều câu hỏi thi yêu cầu chỉ ra các đặc trưng của xử lý tập trung và xử lý phân tán. Ví dụ, các câu hỏi có thể có dạng "Cái nào dưới đây là một đặc trưng thích hợp của hệ thống xử lý tập trung?" Cần biết các ưu điểm và nhược điểm của mỗi kiểu xử lý.

| | |
|---|---|
| Việc bảo trì dễ dàng.<br>Có tính kinh tế vì chỉ những khối cần thiết được cài đặt. | Các khối riêng biệt được quản lý lỏng lẻo. |

Như trong bảng sau đây, hệ thống phân tán có thể được phân chia theo trạng thái phân tán của các chức năng và các tải. Người ta cho rằng phân tán tải theo chiều dọc là không tồn tại trong thực tế.

| Cấu hình<br>Chức năng | Phân tán chức năng | Phân tán tải |
|---|---|---|
| Phân tán theo chiều ngang | Phân tán chức năng theo chiều ngang | Phân tán tải theo chiều ngang |
| Phân tán theo chiều dọc | Phân tán chức năng theo chiều dọc | |

### Phân tán chức năng theo chiều ngang

Một hệ thống phân tán chức năng theo chiều ngang là một hệ thống trong đó các máy tính được phân loại theo kiểu ứng dụng và kiểu dữ liệu; các ví dụ loại này bao gồm sự phân tán chức năng xử lý và sự phân tán cơ sở dữ liệu. Ví dụ, trong các thể chế tài chính, các máy chủ có thể được phân chia thành nhóm trong một hệ thống thông tin và nhóm trong một hệ thống tài khoản; sự phân chia này dựa trên kiểu xử lý, do đó nó là một ví dụ của sự phân tán chức năng xử lý. Sự phân tán cơ sở dữ liệu có nghĩa là các máy tính đó được đặt trong các vị trí thích hợp dựa trên nội dung của dữ liệu.

### Phân tán tải theo chiều ngang

Đây là một hệ thống trong đó nhiều máy tính thực hiện chung các xử lý khi một ứng dụng được thực thi. Khi một xử lý được yêu cầu, một máy tính rỗi sẽ phản hồi. Trong chế độ này, nếu một máy tính bị lỗi, xử lý đó được chuyển đến một máy tính khác và tiếp tục thực hiện. Do đó, hệ thống vẫn hiệu quả trong thời gian bị lỗi. Một hệ thống đa bộ xử lý kết đôi chặt chẽ là một ví dụ của kiểu phân tán này.

### Phân tán chức năng theo chiều dọc

Đây là một hệ thống trong đó chức năng xử lý được chia sẻ giữa các máy trạm dọc theo những người sử dụng cá nhân cũng như các máy tính được chia sẻ bởi nhiều người sử dụng. Do đó, đây là một mối quan hệ theo chiều dọc liên quan đến chức năng xử lý. Một hệ thống khách/chủ là một ví dụ thông thường của hệ thống phân tán chức năng theo chiều dọc.[95]

[95] (Gợi ý) Một hệ thống khách/chủ được đưa ra như thể là một hệ thống phân tán theo chiều ngang, nhưng một cách đúng đắn nó được phân loại là phân tán chức năng theo chiều dọc. Vì một máy chủ thực hiện xử lý của nhiều máy khách, nên đây là một mối quan hệ chức năng theo chiều dọc.

## 2.3.4 Phân loại theo chế độ xử lý

| Điểm chính | ➢ Trong chế độ xử lý theo lô, tất cả các xử lý được lưu trữ và xử lý trong cùng một lần.<br>➢ Trong xử lý thời gian thực, dữ liệu được xử lý ngay khi nó được tạo ra. |
|---|---|

Dựa trên quan điểm về chế độ xử lý, các cấu hình hệ thống có thể được phân thành 2 loại: xử lý theo khối và xử lý thời gian thực. Chúng cũng có thể được phân loại trên cơ sở có hay không kết nối đến một mạng.

| Chế độ xử lý | Chế độ hoạt động | Phương pháp kết nối |
|---|---|---|
| Xử lý theo khối[96] | Xử lý theo khối trung tâm | Ngoại tuyến |
| | Xử lý theo khối từ xa | Trực tuyến |
| Xử lý thời gian thực | Xử lý chế độ tương tác | |
| | Xử lý giao dịch trực tuyến | |
| | Điều khiển thời gian thực | |

### Các hệ thống xử lý theo khối (Các hệ thống xử lý một lần)

Thuật ngữ "khối" (batch) có nghĩa là một gói gộp chung lại (bundle). Xử lý theo khối là bất kỳ một phương pháp nào trong đó dữ liệu để xử lý bởi một máy tính được lưu trữ trong một khoảng thời gian nào đó hoặc được lưu trong một mảng hoặc phương tiện lưu trữ và sau đó được xử lý chung trong cùng một lần. Các hệ thống như vậy có các đặc trưng sau:

- Máy tính có thể được sử dụng hiệu quả vì việc xử lý thực hiện trong một lần.
- Phù hợp với các thủ tục con và các xử lý có tính lặp đi lặp lại (các tác vụ chuẩn).
- Kết quả không đạt được ngay lập tức vì việc xử lý được hoàn thành chung.

Trong các hệ thống xử lý theo khối trung tâm, xử lý khối đưa ra tại một máy tính trung tâm; trong chế độ xử lý theo khối từ xa[97], xử lý khối được thực hiện từ một thiết bị cuối có thể kết nối tới qua một đường truyền thông.

[96] **Xử lý theo khối và xử lý thời gian thực**: Xử lý theo khối là trong đó dữ liệu được lưu trữ và được xử lý chung một lần. Các ví dụ có thể kể tới là tính toán tiền điện, tiền nước, tiền ga. Tính toán bảng lương cũng là một ví dụ của xử lý theo lô. Xử lý thời gian thực là trong đó dữ liệu được xử lý ngay khi nó được sinh ra. Các hệ thống đặt chỗ máy bay, tàu hỏa là các ví dụ của xử lý loại này.

[97] **Xử lý khối trung tâm và xử lý khối từ xa**: Xử lý theo khối trung tâm không sử dụng đường truyền thông. Xử lý theo khối từ xa là trong đó một thiết bị cuối ngay gần được sử dụng để thực hiện xử lý theo khối tại một máy chủ từ một vị trí ở xa. Đôi khi được gọi là RJE, viết tắt của "Remote Job Entry" khi người sử dụng thực hiện một công việc ở vị trí từ xa.

## Các hệ thống xử lý tương tác

Đây là một phương pháp trong đó xử lý được thực hiện trong chế độ hội thoại với máy tính. Các xử lý trong giao tiếp với máy tính có thể sửa chữa các lỗi ngay lập tức mỗi khi chúng được thông báo; vì thế, đây là một chế độ xử lý phù hợp cho việc phát triển chương trình, v.v... Một hệ thống xử lý tương tác yêu cầu các phản hồi nhanh chóng.

Một hệ thống xử lý tương tác cũng yêu cầu một hệ điều hành trong đó hỗ trợ chức năng đa lập trình, và xử lý được thực hiện sử dụng TSS.[98]

## Các hệ thống xử lý giao dịch trực tuyến

Một giao dịch là một chuỗi dữ liệu hoặc lệnh, thường được gọi là một tệp giao dịch hoặc dữ liệu giao dịch. Xử lý giao dịch trực tuyến là xử lý dữ liệu cập nhật trên một tệp trực tuyến tại một máy tính trung tâm thông qua một thiết bị cuối kết nối trực tuyến. Các ví dụ của xử lý giao dịch trực tuyến bao gồm các hệ thống lưu trữ tài khoản trong ngân hàng, các hệ thống đặt chỗ hàng không và đường sắt.

Xử lý giao dịch trực tuyến yêu cầu các thiết bị cuối chuyên dụng có tính hữu dụng cao mục đích để nâng cao hiệu năng xử lý như các ATM ngân hàng, các thiết bị cuối phục vụ phát hành vé tại các cửa đặt vé. Ngoài ra, vì dữ liệu được chia sẻ bởi nhiều thiết bị cuối, có một rủi ro là việc truy cập cùng lúc đến cùng một dữ liệu có thể gây ra vấn đề bế tắc (deadlock) [99] hoặc phá hủy dữ liệu. Vì thế, cần phải nâng cao bảo trì dữ liệu.

## Điều khiển thời gian thực

Nhìn chung, điều khiển thời gian thực là phương pháp trong đó dữ liệu được xử lý trong thời gian thực mỗi khi có một yêu cầu xử lý và kết quả trả về ngay lập tức cho đối tượng yêu cầu. Khái niệm này cũng bao gồm xử lý giao dịch trực tuyến. Tuy nhiên, trong một ngữ cảnh hẹp, đây là chế độ xử lý đặc trưng đặt tại các thiết bị sản xuất, trong đó hệ thống được kết nối với một hệ thống cảm biến các chuyển động vật lý của các đối tượng để điều khiển, các xử lý tương ứng với các tín hiệu bên ngoài được thực thi ngay lập tức, và các kết quả cũng ngay lập tức được gửi trả về cho các khối trên các dây truyền sản xuất như là các tín hiệu điều khiển. Các hệ thống điều khiển sản xuất trong một dây chuyền sản xuất thép và nhà máy ô tô là các ví dụ về hệ thống điều khiển thời gian thực. Một ví dụ khác là một qui trình giám sát hệ thống điện 24/24 giờ quản lý bởi các máy tính; nếu có lỗi xảy ra, hệ thống sẽ thông báo đến cơ quan bảo trì theo thời gian thực, hoặc khối mà dò ra sự cố sẽ bị ngắt theo thời gian thực.[100]

[98] **TSS (Time Sharing System)**: Hệ thống cho phép nhiều chương trình được thực thi theo một thứ tự xác định trong một khoảng thời gian cực ngắn trong mỗi lần. Thủ tục này được lặp lại để việc thực hiện của mỗi chương trình có thể được hoàn thành trong một khoảng thời gian nào đó. Trên quan điểm của người sử dụng, không có sự can thiệp của con người, vì thế người sử dụng có thể cảm thấy như thể là hệ thống máy tính chỉ sử dụng dành riêng cho một chương trình.

[99] **Bế tắc** (**Deadlock)**: Là trường hợp trong đó hệ thống bị tắc nghẽn do nhiều tác vụ (chương trình) cố gắng truy cập vào cùng một tài nguyên (tệp, cơ sở dữ liệu, v.v...) và đi vào trạng thái chờ.

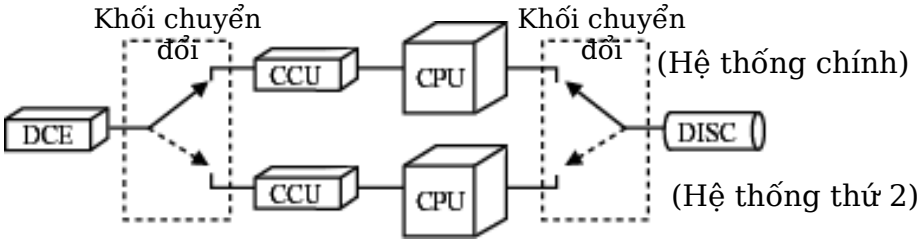
[100] (FAQ) Nhiều câu hỏi thi liên quan đến các đặc trưng của xử lý theo khối từ xa và xử lý tương

tác. Cả 2 đều là xử lý trực tuyến nhưng chú ý rằng xử lý theo khối từ xa là kiểu xử lý theo khối trong khi xử lý tương tác là kiểu xử lý thời gian thực.

## Câu hỏi nhanh

**Q1** Giải thích các vai trò của các máy khách và máy chủ trong một hệ thống khách/chủ.

**Q2** Hình dưới đây chỉ ra loại cấu hình hệ thống nào?




**Q3** Trong bảng dưới, xử lý phân tán được phân loại theo trạng thái phân tán của chức năng và tải. Hệ thống khách chủ thuộc loại nào ?

| Cấu hình / Chức năng | Phân tán chức năng | Phân tán tải |
|---|---|---|
| Phân tán theo chiều ngang | Phân tán chức năng theo chiều ngang | Phân tán tải theo chiều ngang |
| Phân tán theo chiều dọc | Phân chức năng theo chiều dọc | |

**Q4** Sự khác nhau giữa xử lý theo khối và xử lý thời gian thực ?

**A1** Máy khách: Gửi các yêu cầu dịch vụ đến máy chủ.
Các máy chủ: Thực hiện các yêu cầu gửi đến từ máy khách và trả về các kết quả xử lý.

**A2** Hệ thống dự phòng: Các điểm để phân biệt là tồn tại 2 CPU và các khối chuyển đổi.

**A3** Phân tán chức năng theo chiều dọc.

**A3** Xử lý theo khối: Dữ liệu được lưu trữ và xử lý trong cùng một lần.

**A4** Xử lý thời gian thực: Dữ liệu được xử lý ngay lập tức khi chúng được tạo ra.

# 2.4 Hiệu năng và sự tin cậy của hệ thống

## Giới thiệu

Để đánh giá các hệ thống máy tính, nhiều phương pháp khác nhau được đưa ra. Trong đó hiệu năng tốt (tốc độ xử lý cao) và khả năng chống lỗi (tính tin cậy cao) là rất quan trọng.

## 2.4.1 Các chỉ số hiệu năng

| Điểm chính | ➢ Thời gian đáp ứng, thông lượng, và thời gian quay vòng được sử dụng để đánh giá hiệu năng.<br>➢ Hỗn hợp lệnh và chỉ số benchmark được sử dụng làm các chỉ số hiệu năng. |
|---|---|

Để đánh giá hiệu năng toàn diện của các hệ thống máy tính, bao gồm phần mềm và phần cứng, có thể sử dụng nhiều tiêu chí khác nhau như thời gian đáp ứng, thông lượng, và thời gian thay đổi. Các chỉ số để đánh giá hiệu năng, đặc biệt là phần cứng bao gồm hỗn hợp lệnh và chỉ số benchmark.

### Các thuật ngữ liên quan đến đánh giá hiệu năng của các hệ thống máy tính

Để đánh giá hiệu năng của các hệ thống máy tinh, việc tính thời gian xử lý của các công việc và các chương trình là quan trọng.[101]

#### Thời gian đáp ứng

Là lượng thời gian từ lúc hoàn thành quá trình nhập từ một đầu vào đến lúc bắt đầu quá trình xuất từ một đầu ra. Ví dụ, khi một yêu cầu xử lý sinh ra từ bàn phím máy tính, thời gian này là lượng thời gian cho đến kết quả được hiển thị trên khối hiện thị hoặc cho đến khi bắt đầu in ra. Đây cũng là yếu tố chính được sử dụng để đánh giá hiệu năng của một hệ thống trực tuyến.

#### Thông lượng (Khả năng xử lý)

Là số lượng công việc có thể được xử lý bởi một hệ thống máy tính trong một đơn vị thời gian nhất định, hoặc là số lượng thời gian để xử lý một công việc nào đó. Thời gian xử lý này bao gồm thời gian CPU dành riêng và thời gian

[101] (Gợi ý) Hỗn hợp lệnh và chỉ số benchmark dùng để đánh giá hiệu năng của các máy tính, hỗn hợp lệnh dùng trong đánh giá phần cứng. Tuy nhiên, mặc dù phần cứng có nhanh thì toàn bộ hiệu năng hệ thống vẫn thấp nếu hiệu năng của hệ điều hành thấp. Do đó, hiệu năng của phần cứng thường chỉ mang tính tham khảo.

đợi xử lý như chuẩn bị các thao tác vào ra và thời gian dọn dẹp.

### Thời gian quay vòng (TAT)

Về mặt kỹ thuật, đây là lượng thời gian để tạo ra sự quay vòng của hệ thống. Trong xử lý theo khối, đây là lượng thời gian tồn tại từ sự xem xét của một chương trình tại cửa sổ đến khi đạt được các kết quả. Trong các hoạt động kinh doanh, đây là lượng thời gian từ khi một một khách hàng đưa ra một yêu cầu đặt hàng đến khi sản phẩm đặt hàng được chuyển tới khách hàng.

## Hỗn hợp lệnh (Instruction Mix)[102]

Hỗn hợp lệnh được sử dụng để so sánh hiệu năng phần cứng trong các hệ thống máy tính. Cho dù phần cứng tốt, nếu hiệu năng của hệ điều hành thấp thì hiệu năng của toàn bộ hệ thống cũng thấp. Một hỗn hợp lệnh dùng cho một chương trình mức độ trung bình để tính toán thời gian thực hiện lệnh trung bình trên một lệnh và giá trị MIPS,[103] dựa trên tần xuất thực hiện của mỗi lệnh.

Sử dụng các điều kiện sau, chúng ta sẽ đưa ra một vài tính toán cụ thể giá trị MIPS

| Nhóm lệnh | Tốc độ thực hiện (microsecond) | Tần số xuất hiện |
|---|---|---|
| A | 0.1 | 40% |
| B | 0.2 | 30% |
| C | 0.5 | 30% |

Đầu tiên, chúng ta tính thời gian thực hiện lệnh trung bình. Tốc độ thực hiện mỗi lệnh được biểu diễn theo đơn vị micro giây ($10^{-6}$). Thời gian thực hiện lệnh trung bình là tổng thời gian (trên toàn bộ các lệnh) của tích số giữa thời gian thực hiện lệnh với tần số xuất hiện tương ứng.

Thời gian thực hiện lệnh trung bình $= 0.1 * 10^{-6} * 0.4 + 0.2 * 10^{-6} * 0.3 + 0.5 * 10^{-6} * 0.3$

$= (0.04 + 0.06 + 0.15) * 10^{-6}$

$= 0.25 * 10^{-6}$ (giây/lệnh)

Số lệnh trung bình thực hiện trong một giây là nghịch đảo của thời gian thực hiện lệnh trung bình, tính như sau:

Số lệnh trung bình thực hiện trong một giây $= 1 / (0.25 * 10^{-6})$

$= 4 * 10^{6}$ (lệnh/giây)

$= 4$ (MIPS).[104] [105]

[102] (Gợi ý) Một hỗn hợp lệnh dành cho các tính toán khoa học gọi là "hỗn hợp Gibson", và một hỗn hợp lệnh khác dành cho các tính toán kinh doanh gọi là "hỗn hợp thương mại".

[103] **MIPS (Million Instructions Per Second)**: Đây là chỉ số hiệu năng miêu tả số lệnh máy trong một giây theo đơn vị triệu ($10^6$). Đây chỉ là hiệu năng của phần cứng, vì vậy nó chỉ được sử dụng mang tính tham khảo.

[104] (FAQ) Các kỳ thi đưa ra các câu hỏi yêu cầu tính giá trị MIPS đưa ra một hỗn hợp lệnh hoặc tính thời gian trung bình của một lệnh. Bạn nên làm quen với các câu hỏi tính toán đó.

[105] **Clock**: Là tần số của tín hiệu xung nhịp được tạo ra bởi bộ tạo xung nhịp. Bởi vì các lệnh bên trong CPU được đồng bộ với tín hiệu xung nhịp khi chúng được thực hiện, nên tần số xung nhịp càng cao, số lệnh được thực hiện trong một khoảng thời gian càng lớn. Ví dụ, nếu tần số xung

FLOPS[106] được sử dụng như một chỉ số để tính hiệu năng của thao tác với các số chấm động.

## Chỉ số Benchmark

Chỉ số benchmark được sử dụng để so sánh và đánh giá hiệu năng toàn diện của các máy tính, bao gồm phần cứng và hệ điều hành, bằng cách đo thời gian thực thi của chương trình chuẩn[107]

# 2.4.2 Tính tin cậy

| **Điểm chính** | ➢ Các chỉ số tin cậy gồm RAS, RASIS, và đường cong bathtub<br>➢ Các điểm cần lưu ý trong thiết kế tính tin cậy là khả năng chịu lỗi và khả năng mềm dẻo với lỗi. |
|---|---|

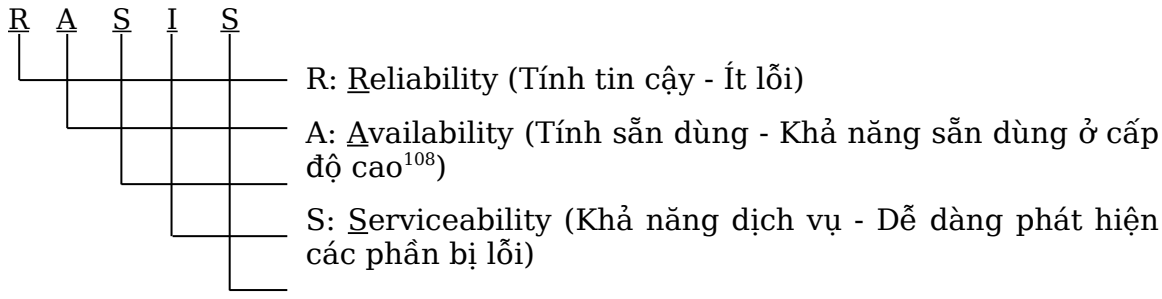
Có nhiều mức độ tin cậy được yêu cầu đối với hệ thống thông tin phụ thuộc vào mục đích sử dụng của hệ thống. Đôi khi phải chấp nhận các yếu tố về kinh tế để đạt được tính tin cậy cao. Trong một vài tình huống khác, không chỉ yếu tố tính tin cậy đối với các hoạt động của hệ thống mà các thông tin được điều khiển bởi hệ thống cũng cần phải đảm bảo tin cậy.

## Các chỉ số tính tin cậy

Tính tin cậy là cấp độ trong đó hệ điều hành hoạt động ổn định. Trường hợp lý tưởng là hệ thống không bị lỗi, nhưng không có hệ thống nào mà không bị bất cứ một lỗi gì.

### RAS/RASIS

Cả 2 thuật ngữ RAS và RASIS đều là viết tắt của những yếu tố giúp cho các hệ thống máy tính hoạt động trong một trạng thái ổn định. RAS là viết tắt 3 chữ cái đầu tiên của RASIS:


R A S I S
R: Reliability (Tính tin cậy - Ít lỗi)
A: Availability (Tính sẵn dùng - Khả năng sẵn dùng ở cấp độ cao[108])
S: Serviceability (Khả năng dịch vụ - Dễ dàng phát hiện các phần bị lỗi)

nhịp là 200MHz, có $200 * 10^6$ tín hiệu xung nhịp trên giây. Thông thường, một lệnh chiếm vài xung nhịp.

[106] **FLOPS**: Viết tắt của floating-point operations per second (Số phép toán dấu phẩy động trong 1 giây). Nếu miêu tả theo đơn vị triệu, kí hiệu là MFLOPS.

[107] (Gợi ý) Một ví dụ về chỉ số benchmark trong tính toán ký thuật là SPECmark, và một ví dụ về chỉ số benchmark giao dịch là TPC. TPC-C là chỉ số benchmark thường được sử dụng trong nhóm TPC, đáp ứng trực tiếp các ứng dụng kinh doanh trên thực tế.

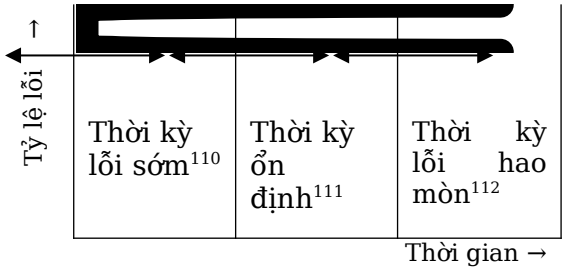
I: Integrity (Tính toàn vẹn - Nhất quán dữ liệu)

S: Security (Sự an toàn - Ngăn cản các truy cập không được phép)

[108] **Availability**: Tính sẵn dùng được biết đến như là khả năng mà hệ thống duy trì các chức năng (hoặc động) tại bất kỳ thời điểm nào hoặc phần trăm lượng thời gian mà các chức năng được duy trì trong suốt một khoảng thời gian nào đó.

### Đường cong Bathtub

Đường cong Bathtub được sử dụng để minh họa cho khái niệm vòng đời phần cứng. Phần cứng có thể bị lỗi trong thời gian đầu của quá trình vận hành do một số bộ phận bị hỏng hóc, nhưng xác suất xảy ra các lỗi khi đó sẽ giảm dần do việc sửa chữa và thay thế. Sau đó, do hao mòn và hỏng hóc của nhiều bộ phận, xác suất xảy ra lỗi lại gia tăng, và thậm chí vòng đời của nó sẽ kết thúc. Đường cong này được chỉ ra như dưới đây.[109]




# Các điểm lưu ý trong thiết kế tính tin cậy

Một hệ thống tin cậy cao có thể tiếp tục vận hành thậm chí ngay cả khi một vài phần của hệ thống bị lỗi, hệ thống như vậy được gọi là một hệ thống chịu lỗi (fault-tolerant). Các kỹ thuật phổ biến cho các cấu hình hệ thống tin cậy cao bao gồm kỹ thuật mềm dẻo với lỗi (fail-soft), chức năng cho phép hệ thống tiếp tục hoạt động của nó với hiệu năng thấp hơn hoặc ít chức năng hơn khi một lỗi xảy ra; và kỹ thuật an toàn với lỗi (fail-safe), chức năng cho phép hệ thống hoạt động an toàn bằng cách tránh các điều kiện rủi ro khi có một lỗi xảy ra.

### Kỹ thuật mềm dẻo với lỗi (Fail-soft)

Đây là chức năng trong đó khi có một lỗi xảy ra, bộ phận bị lỗi được ngắt khỏi hệ thống và hệ thống tiếp tục hoạt động với một mức độ hiệu năng thấp hơn (fall back[113]). Trong một hệ thống dự phòng, thông thường có 2 hệ thống cùng xử lý dữ liệu độc lập, nhưng nếu một hệ thống bị lỗi, cấu hình sẽ chuyển xử lý sang hệ thống còn lại và sẽ tiếp tục xử lý. Ngoài ra, khi một lỗi xảy ra trong hệ đa bộ xử lý, hệ thống tiếp tục các dịch vụ bằng cách ngắt bộ xử lý bị hỏng ra khỏi hệ thống. Đây cũng là một cấu hình hệ thống mềm dẻo với lỗi.

### Kỹ thuật an toàn với lỗi (Fail-safe)

Đây là một chức năng trong đó khi có một lỗi xảy ra, hệ thống sẽ khóa các chức năng của nó trong một chế độ an toàn đã thiết lập để chuyển đến điều khiển khu vực ảnh hưởng của lỗi[114]. Điều này cũng giống như là phương pháp

[109] (Gợi ý) Đường cong bathtub được đặt tên như vậy là vì đồ thị chỉ ra mối quan hệ giữa xác suất lỗi và thời gian giống với hình dạng của cái bồn tắm (bath-tub).
[110] **Thời kỳ lỗi sớm** (**Early failure period)**: Thời kỳ của các lỗi tại thời điểm bắt đầu sử dụng của các khối. Các lỗi sẽ xuất hiện ít hơn theo thời gian.
[111] **Thời kỳ ổn định (Stable failure period)**: Các khối ổn định trong suốt thời kỳ này, ít xuất hiện các lỗi.
[112] **Thời kỳ lỗi hao mòn (Wear-out failure period)**: Một khoảng thời gian nào đó trôi qua, các lỗi trở nên xuất hiện nhiều hơn trong thời kỳ này.
[113] **Fall-back**: Trong một hệ thống máy tính mềm dẻo với lỗi, việc xử lý vẫn tiếp tục với các chức năng ở mức độ thấp hơn; điều này gọi là fall-back.
[114] (Gợi ý) Mềm dẻo với lỗi (Fail-sofe) và an toàn với lỗi (fail-safe) là các từ tương tự nhau, tránh nhầm lẫn chúng.

tất cả các đèn báo của đường sắt đều bật đỏ khi có một tai nạn xảy ra. Trong các cấu hình hệ thống mà có 2 hệ thống so sánh các kết quả xử lý với nhau, như trong một hệ thống song hành, khi các kết quả so sánh khác nhau, hệ thống mà được xác định xảy ra lỗi sẽ bị ngắt trong khi hoạt động vẫn được tiếp tục.

**Fool-proof**

Thuật ngữ này là một phương pháp ngăn chặn một công dụng không định trước của một chương trình gây lỗi, đặc biệt khi nhiều người sử dụng cùng sử dụng một chương trình. Nếu một cá nhân đang sử dụng một chương trình cụ thể, cách mà chương trình thực hiện không gây ra vấn đề gì nhưng khi có nhiều người cùng sử dụng, chương trình được sử dụng như thế nào sẽ rất khó dự đoán.[115]

## 2.4.3 Tính sẵn dùng

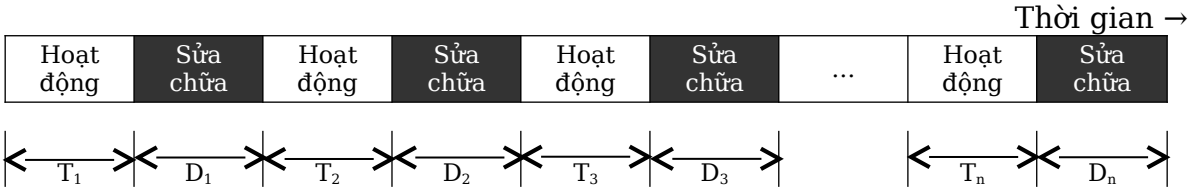
| **Điểm chính** | ➢ MTBF là thời gian hệ thống hoạt động đúng đắn, và MTTR là thời gian hệ thống được sửa chữa.<br>➢ Tính sẵn dùng là tỷ lệ thời gian khi hệ thống hoạt động đúng đắn. |
|---|---|

Một trong những chỉ số trong RASIS là "A" viết tắt của Availability, có nghĩa là tỷ lệ hoạt động, được tính sử dụng MTBF và MTTR như sau:

Tính sẵn dùng: A = (MTBF) / (MTBF + MTTR)

## MTBF và MTTR

Giả sử trạng thái hoạt động của một hệ thống máy tính được chỉ ra như trong hình sau:




**MTBF (Thời gian trung bình giữa các lỗi - Mean Time Between Failures)**

Đây là thời gian trung bình mà hệ thống tiếp tục hoạt động không có lỗi. Giá trị MTBF càng lớn, hệ thống càng tin cậy. Vì thế, nó được sử dụng như một chỉ số để đánh giá tính tin cậy ("R" trong RASIS) [116]

$MTBF = (T_1 + T_2 + T_3 + \ldots T_n) / n$

(Trong đó, "n" là số khoảng thời gian hệ thống hoạt động không có lỗi.)

[115] (FAQ) Có những câu hỏi thi liên quan đến ý nghĩa của các chữ cái RASIS là gì khi nó là một chỉ số tính tin cậy của hệ thống máy tính.

[116] (Gợi ý) Các chức năng cải tiến MTBF gồm phát hiện lỗi, tự động sửa lỗi 1 bit, thử lại lệnh, v.v… Đó là các chức năng ngăn chặn hệ thống không bị dừng. Các chức năng cải tiến MTTR gồm đưa ra nhật ký hoạt động (log). Bằng cách tìm các nhật ký hoạt động (logs), có thể xác định được nguyên nhân của lỗi. Việc bảo trì từ xa cũng giúp phát hiện lỗi nhanh chóng, cải tiến MTTR.

### MTTR (Thời gian trung bình để sửa chữa - Mean Time To Repair)

Đây là khoảng thời gian trung bình cần để sửa chữa khi một lỗi xảy ra. Thời gian sửa chữa càng ngắn, hệ thống càng tốt. Vì thế, nó là một chỉ số đánh giá khả năng dịch vụ (“S” trong RASIS).

$MTTR = (D_1 + D_2 + D_3 + \dots D_n) / n$

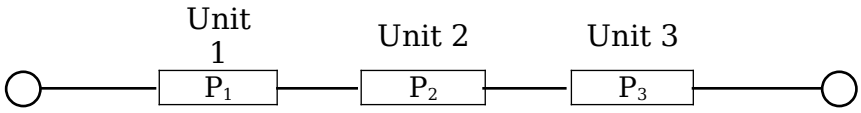
(Trong đó n là số khoảng thời gian hệ thống hoạt động không có lỗi.)

## Đánh giá tính sẵn dùng

Để tính toán tính sẵn dùng, các kết nối tuần tự và kết nối song song phải được tính toán khác nhau. Ý tưởng cơ bản được mô tả như sau[117]

### Tính sẵn dùng trong một hệ thống kết nối tuần tự

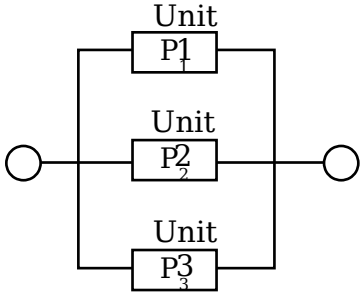
Tính sẵn dùng của toàn bộ hệ thống kết nối tuần tự được chỉ ra ở đây là tích số tính sẵn dùng của mỗi khối. Ở đây, $P_1$, $P_2$, và $P_3$ là tính sẵn dùng của các khối tương ứng như trong hình vẽ.




Tính sẵn dùng của toàn bộ hệ thống = $P_1 * P_2 * P_3$

### Tính sẵn dùng trong một hệ thống kết nối song song

Giả sử rằng có một hệ thống kết nối song song như hình dưới đây, trong đó hệ thống hoạt động miễn là ít nhất một trong các khối 1, 2, 3 hoạt động. Ở đây, tín sẵn dùng được tính toán sử dụng lập luận rằng xác suất của toàn bộ không gian mẫu là 1.



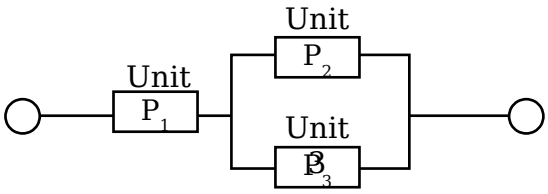

Tính sẵn dùng của toàn bộ hệ thống
= 1 – (Xác suất tất cả các khối cùng hỏng đồng thời)
= 1 – (Xác suất khối 1 hỏng) * (Xác suất khối 2 hỏng) * (Xác suất khối 3 hỏng)

[117] (FAQ) Đây luôn là một câu hỏi liên quan đến tính toán sẵn dùng. Bạn cần chắc chắn hiểu cách tính nó.

$$= 1 - (1 - P_1) * (1 - P_2) * (1 - P_3)$$

**Tính sẵn dùng trong một hệ thống kết hợp cả kết nối tuần tự và song song**

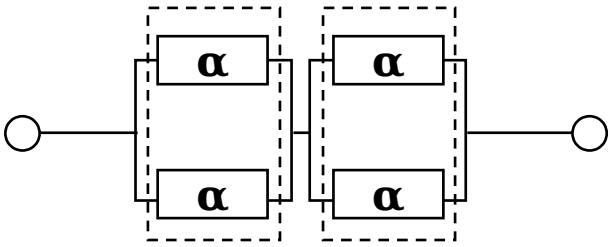
Giả sử rằng có một hệ thống hoạt động nếu khối 1 hoạt động và ít nhất một trong số các khối 2 và 3 hoạt động. Trong trường hợp này, khối 1 và cụm khối 2 và 3 là kết nối tuần tự.




Tính sẵn dùng
= (Tính sẵn dùng của khối 1) ∗ {1 – (Xác suất khối 2 và 3 cùng hỏng đồng thời)}
= (Tính sẵn dùng của khối 1) ∗ {1 – (Xác suất khối 2 hỏng) * (Xác suất khối 3 hỏng)}
= $P_1 * \{1 - (1 - P_2) * (1 - P_3)\}$

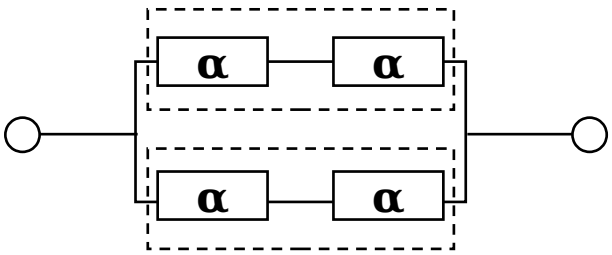
Xem xét các cấu hình phức tạp hơn[118]. Mặc dù 2 hệ thống dưới đây có thể xuất hiện giống nhau nhưng tính sẵn dùng của chúng là khác nhau. Ở đây, tính sẵn dùng được biểu diễn bằng chữ α trong hình.

[Cấu hình 1] Có 2 cụm kết nối song song (bên trong đường nét đứt) được kết nối tuần tự.




Tính sẵn dùng = $\{1 - (1 - \alpha)^2\} * \{1 - (1 - \alpha)^2\} = \alpha^2 (2 - \alpha)^2$

[Cấu hình 2] Có 2 khối kết nối nối tiếp (bên trong đường nét đứt) được kết nối song song.




Tính sẵn dùng = $1-(1 - \alpha^2) * (1 - \alpha^2) = \alpha^2 (2 - \alpha^2)$

[118] (Gợi ý) Chú ý rằng các cấu hình tương tự nhau có tính sẵn dùng khác nhau.

## Câu hỏi nhanh

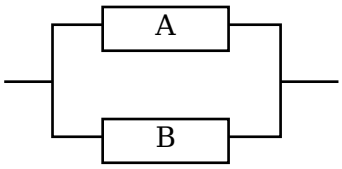
**Q1** Giải thích ý nghĩa của các thuật ngữ: “MIPS,” “thời gian đáp ứng,” “thông lượng,” và thời gian xoay vòng (“turn-around time.”)

**Q2** Giải thích ý nghĩa của RASIS.

**Q3** Giải thích ý nghĩa của MTBF và MTTR.

**Q4** Biểu diễn tính sẵn dụng sử dụng MTBF và MTTR.

**Q5** Tính tính sẵn dùng của toàn bộ hệ thống chỉ ra như hình dưới. A và B là 2 khối, mỗi khối có tính sẵn dùng là 0.97. Toàn bộ hệ thống được xem là hoạt động nếu ít nhất một trong 2 khối hoạt động.




**A1**

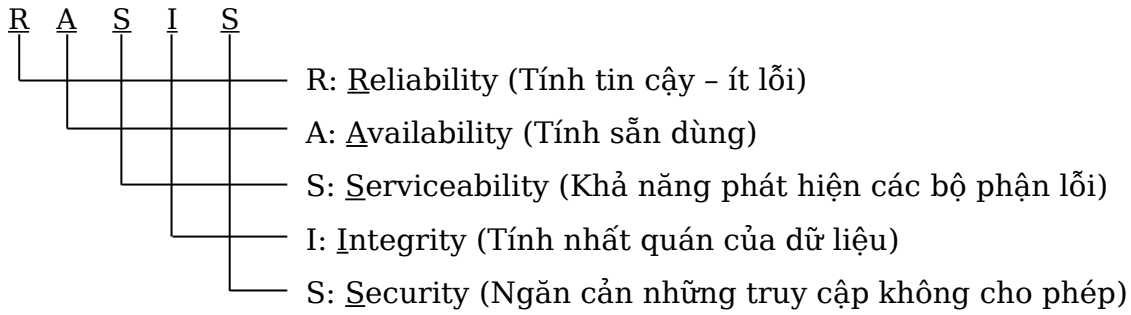
MIPS: Viết tắt của “million instructions per second.” Chỉ ra số lệnh có thể được thực hiện trong 1 giây, biểu diễn theo đơn vị hàng triệu ($10^6$). Đây là một chỉ số hiệu năng phần cứng.

Thời gian đáp ứng: Là lượng thời gian trong khoảng từ khi truyền dữ liệu đi đến khi trả về kết quả xử lý. Đây là một chỉ số hiệu năng của một hệ thống trực tuyến.

Thông lượng: Là một phương pháp đánh giá khả năng xử lý công việc của một hệ thống trong một lượng thời gian. Đây là một trong những chỉ số đánh giá hiệu năng của hệ thống máy tính.

Thời gian xoay vòng: Là lượng thời gian trong khoảng từ khi có một yêu cầu công việc đến khi hoàn thành đầu ra. Chủ yếu được sử dụng trong xử lý theo khối, đây là một chỉ số đánh giá hệ thống bao gồm hoạt động của nó.

**A2** RASIS là viết tắt của các yếu tố giúp cho các hệ thống máy tính vận hành trong một trạng thái ổn định.


R A S I S
R: Reliability (Tính tin cậy – ít lỗi)
A: Availability (Tính sẵn dùng)
S: Serviceability (Khả năng phát hiện các bộ phận lỗi)
I: Integrity (Tính nhất quán của dữ liệu)
S: Security (Ngăn cản những truy cập không cho phép)

**A3** MTBF: (Mean Time Between Failures) là lượng thời gian trung bình mà hệ thống tiếp tục hoạt động không có lỗi.
MTTR: (Mean Time To Repair) là lượng thời gian trung bình để sửa chữa khi có lỗi xảy ra.

**A4** Availability (Tính sẵn dùng): A = (MTBF) / (MTBF + MTTR)

**A5** Chú ý rằng 2 khối được kết nối song song.

$$\begin{aligned}\text{Availability} &= 1 - (1 - 0.97) * (1 - 0.97) \\ &= 1 - 0.03 * 0.03 \\ &= 1 - 0.0009 \\ &= 0.9991.\end{aligned}$$

# 2.5 Ứng dụng hệ thống

## Giới thiệu

Hiện nay có khá nhiều hệ thống khác nhau đã được phát triển dựa trên cơ sở dữ liệu và mạng máy tính. Dễ thấy nhất là Internet và các dịch vụ cơ sở dữ liệu (thường được gọi là cơ sở dữ liệu thương mại). Đồ họa 3 chiều cũng là một thí dụ về ứng dụng của hệ thống đa phương tiện.

## 2.5.1 Các ứng dụng mạng

| Điểm chính | ➢ Sử dụng hạ tầng Web, Internet, mạng nội bộ, mạng nội bộ mở rộng.<br>➢ Các hệ thống ứng dụng bao gồm mua bán qua internet, phần mềm nhóm và thẻ tín dụng. |
|---|---|

Ngày nay, chúng ta đã có một hệ thống mạng thông tin khổng lồ trải rộng khắp thế giới. Có rất nhiều hệ thống sử dụng mạng này và những ứng dụng trên đó tương đối phong phú.
Đây là lĩnh vực khá mới, do vậy sẽ không có nhiều câu hỏi trong chủ đề này, chúng cũng khá dễ. Phần lớn câu hỏi sẽ chỉ yêu cầu kiến thức ở mức khái niệm, vì vậy cần nhớ những khái niệm cơ bản để đạt kết quả tốt nhất.

## Các ứng dụng cơ sở hạ tầng

"Cơ sở hạ tầng" có nghĩa là "cơ sở" hoặc "nền tảng." Trong các hệ thống máy tính, từ này ám chỉ đến các thành phần cơ sở bao gồm phần những mềm và phần cứng cấu thành nên hệ thống đó. Vi dụ, trong một cấu trúc mạng, các thành phần khác nhau như dây cáp, các thiết bị truyền thông tin, và hệ thống quản lý các dây dẫn là các bộ phận của một **cơ sở hạ tầng truyền thông**.

**Web (WWW: World Wide Web):**Hệ thống cup cấp thông tin ở dạng siêu văn bản, được phát triển bởi các nhà nghiên cứu ở CERN. Thông tin phân bố trên khắp thế giới được liên kết lại nhờ sử dụng siêu văn bản và mạng Internet.

WWW là cơ chế lưu trữ thông tin dưới dạng một trang chủ trên mỗi máy chủ phục vụ. Phần mềm truy cập WWW và hiển thị lên màn hình gọi là Web browser hoặc đơn giản trình duyệt web - "browser".

**Internet**

Là một tập hợp của các mạng trên toàn thế giới kết nối với nhau thông qua giao thức TCP/IP. Không có một tổ chức chính phủ nào hoặc một tổ chức được chỉ định nào quản lý tập trung toàn bộ Internet. Thay vào đó, việc hỗ trợ kỹ thuật và quản lý tài nguyên trên Internet được thực hiện bởi các tổ chức tình

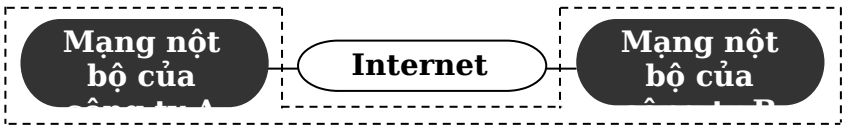
nguyện. Mạng ARPANET[119], được thành lập bởi bộ quốc phòng Mỹ, chính là tiền thân của Internet.

### Intranets – Mạng nội bộ

Mạng nội bộ là mạng trong phạm vi của một công ty ứng dụng công nghệ của Internet. Thông thường một bức tường lửa[120] sẽ dược thiết lập giữa Internet và mạng để ngăn chặn rò rỉ thông tin của công ty ra ngoài.

### Extranets – Mạng nội bộ mở rộng

Extranet là một mạng intranet mở rộng của nhiều công ty. Nói chung, extranet được xây dựng bởi việc kết nối nhiều intranet vào Internet




### Truyền thông di động

**Truyền thông di động**[121] là môi trường mạng mà người dùng có thể truy cập được từ bất cứ nơi nào. Ngày nay truyền thông là xu hướng chủ đạo, chúng ta có thể gửi và nhận thư điện tử trên Internet và tìm kiếm được nhiều thông tin đa dạng chỉ với một chiếc điện thoại.

### Liên lạc vệ tinh

**Liên lạc vệ tinh** là một hệ thống truyền dẫn không dây sử dụng các vệ tinh truyền thông. Một trạm phát truyền đi (đường lên) một lượng lớn thông tin đến một vệ tinh truyền thông cách xa khoảng 36000km trên quỹ đạo trái đất, thông tin sau đó được phát trở lại (downlink) tới nhiều trạm thu trên trái đất. Bằng cách đó, một lượng lớn thông tin có thể truyền tới nhiều điểm cùng một lúc.

### CATV

Nhiều loại hình dịch vụ sử dụng CATV (truyền hình cáp) đang được xem xét và thảo luận để đưa vào thương mai hóa, như là kết nối Internet, dịch vụ điện thoại, các thử nghiệm liên quan đến PHS (Personal Handy-phone System), và VOD ( Video theo yêu cầu). CATV được mong đợi sẽ là thành phần chính của hạ tầng thông tin trong kỷ nguyên đa phương tiện.

---

[119] **ARPANET (Advanced Research Project Agency Network)**: Đây là mạng máy tính quốc gia được phát triển dưới sự tài trợ của Advanced Reseach Project Agency Network ( Dự án mạng hỗ trợ nghiên cứu cao cấp) của bộ quốc phòng Mỹ..

[120] **Tường lửa** - **Firewall**: Là một thiết bị đặt giữa Internet và mạng nội bộ công ty để quản lý dữ liệu vào ra và bảo vệ mạng nội bộ khỏi những tấn công từ bên ngoài và những truy cập không hợp lệ. Bản thân tên của thiết bị cũng bao hàm chức năng của thiết bị.

[121] **Thiết bị di động/Điện toán di động** - **Mobile/mobile computing**: “Thiết bị di động" ám chỉ bất kỳ thiết bị thông tin nào có thể cầm đi được bao gồm điện thoại di động, hệ thống điện thoại cầm tay cá nhân (PHS), máy tính xách tay...“"Điện toán di động” ám chỉ đến chế độ sử dụng của các thiết thông tin để truy cập vào mạng công ty từ xa.

# Các hệ thống ứng dụng

Một hệ thống ứng dụng mạng là tập hợp các hệ thống sử dụng mạng thông tin, bao gồm các ứng dụng sau:

**Mua sắm qua Internet - Internet shopping**
Đây là hệ thống mà người dùng có thể mua hàng ở các cửa hàng ảo trên các trang web. Để thực hiện thanh toán, người mua hàng có thể sử dụng thẻ tín dụng hoặc đi đến một cửa hàng gần đó để trả.
**Phần mềm nhóm - Groupware**

Đây là phần mềm để liên lạc bên trong một tổ chức hoặc để chia sẻ thông tin. Một vài chức năng của phần mềm nhóm: thư điện tử, chia sẽ lịch làm việc, tài liệu và dòng công việc.

**Thẻ tín dụng - Debit cards**

Đây là dịch vụ cho phép người dùng có thể thực hiện thanh toán thông qua một chiếc thẻ tín dụng được ngân hàng phát hành. Tiền thanh toán sẽ được trừ ngay vào tài khoản ngân hàng.

## 2.5.2 Các ứng dụng cơ sở dữ liệu

| Điểm chính | ➢ Thí dụ về một ứng dụng cơ sở dữ liệu – data warehouse.<br>➢ Ứng dụng của cơ sở dữ liệu trong thương mại bao gồm kế toán doanh nghiệp, quản lý tài sản, quản lý tài liệu và hỗ trợ bán hàng. |
|---|---|

Một kiểu hệ thống cơ sở dữ liệu ứng dụng là trung tâm dữ liệu (Data Warehouse). Hệ thống ứng dụng mà cơ sở dữ liệu được sử dụng trong thương mại bao gồm hệ thống kế toán doanh nghiệp[122], hệ thống quản lý tài sản,[123], hệ thống quản lý tài liệu[124], và hệ thống hỗ trợ bán hàng[125].

### Trung tâm dữ liệu - Data Warehouse

Trung tâm dữ liệu là một cơ sở dữ liệu lớn của toàn công ty hỗ trợ trong việc ra quyết định. Ý tưởng của nó là tổ chức lưu trữ một lượng lớn dữ liệu sử dụng trợ giúp trong việc ra quyết định kinh doanh. Đôi khi trung tâm dữ liệu được gọi là cơ sở dữ liệu thông tin (informational database).

### Khai phá dữ liệu - Data Mining

Thuật ngữ này liên quan đến một công nghệ, phương pháp phác thảo ra các xu hướng, sự tương quan, các mô hình dữ liệu cần cho quản lý và tiếp thị, thông qua việc nghiên cứu một lượng lớn dữ liệu thô.

Trung tâm dữ liệu thường phân tích các dữ liệu khác nhau dựa trên một vài giả thuyết còn khai phá dữ liệu thường phát hiện ra những xu hướng và mô hình để từ đó lập ra giả thuyết.

### Chợ dữ liệu - Data Mart

Chợ dữ liệu là cơ sở dữ liệu lưu trữ dữ liệu được trích ra từ trung tâm dữ liệu. Dữ liệu lưu trong này được lựa chọn và thống kê tùy theo mục đích

[122] **Hệ thống kế toán doanh nghiệp** - **Corporate accounting system**: Là hệ thống mà các thủ tục kế toán của công ty được tự động hóa làm cho việc tính toán thống kê hiệu quả và nhanh chóng hơn.
[123] **Hệ thống quản lý tài sản** - **Inventory management system**: Là hệ thống theo dõi cân đối sản xuất và nhu cầu sản phẩm, quản lý tài sản như sản phẩm, nguyên liệu của công ty ở mức tối ưu. Trong các cửa hàng bán lẻ như siêu thị , thông tin bán lẻ nhập vào ở thiết bị POS được thu thập lại và phân tích để sau đó dự đoán nhu cầu và đặt hàng tự động, đảm bảo đủ hàng ở mức an toàn và tối ưu khi mua thêm.
[124] **Hệ thống quản lý tài liệu** - **Document management system**: Đó là hệ thống hỗ trợ công ty trong việc quản lý nhiều loại tài liệu; tìm kiếm tài liệu từ nhiều nguồn lưu trữ khác nhau hoặc từ nội dung của tài liệu . Mục tiêu của hệ thống là làm cho việc quản lý tài liệu và chuẩn bị tài liệu hiệu quả hơn, thí dụ : tránh được việc chuẩn bị lại một tài liệu đã có sẵn trong hệ thống.
[125] **Hệ thống hỗ trợ bán lẻ** - **Sales support system**: Đó là hệ thống hỗ trợ các kế hoạch bán lẻ sản phẩm dựa trên các thông tin bán lẻ tích lũy được.

của từng nhóm người sử dụng.
Khác với trung tâm dữ liệu, nơi chứa toàn bộ thông tin của công ty, chợ dữ liệu có khá ít thông tin đủ để đáp ứng nhu cầu của người sử dụng đích.

## OLAP

OLAP (Xử lý phân tích trực tuyến - OnLine Analytical Processing) là ứng dụng phân tích trong đó người dùng cuối phát hiện ra các vấn đề và giải pháp của vấn đề đó bằng cách trực tiếp tìm kiếm và tổ chức một cơ sở dữ liệu; mục đích là nhanh chóng truy cập tới dữ liệu và cung cấp một chức năng hỗ trợ cho việc phân tích dễ dàng hơn[126].

## OLTP

OLTP (Xử lý giao dịch trực tuyến - OnLine Transaction Process) là chế độ xử lý mà thông điệp được gửi đến máy tính chủ xử lý từ nhiều thiết bị cuối, thông điệp sẽ được xử lý thông qua việc truy cập đến hàng loạt cơ sở dữ liệu và trả lại kết quả ngay lập tức.

Cơ sở dữ liệu sử dụng bởi OLTP được gọi là cơ sở dữ liệu doanh nghiệp (business databases), hoặc đôi khi cơ sở dữ liệu trực tuyến (operational databases). Những khái niệm này khác với cơ sở dữ liệu thông tin (informational databases)

Bảng dưới đây so sánh hai loại cơ sở dữ liệu trên:

| Tiêu chí so sánh | Cơ sở dữ liệu thông tin - Informational databases | Cơ sở dữ liệu trực tuyến- Operational databases |
|---|---|---|
| Tác vụ | Trợ giúp quyết định | Các tác vụ doanh nghiệp chung |
| Thêm dữ liệu ? | Có | Có |
| Cập nhật dữ liệu ? | Thường không hỗ trợ | Thường có hỗ trợ |
| Chế độ xử lý | OLAP | OLTP |
| Người sử dụng chính | Người quản lý | Nhân viên |
| Loại hình thương mại | Các tác vụ không theo chẩn[127] | Các tác vụ chuẩn |
| Thời gian duy trì dữ liệu | Dài | Ngắn |

# Các hệ thống ứng dụng

Nhiều hệ thống đã được phát triển dựa trên cơ sở dữ liệu. Ngày nay, sẽ không là quá đáng nếu nói phần lớn các hệ thống đều hoạt động trên cơ sở dữ liệu. Có thể kể ra một vài ưu điểm của việc sử dụng cơ sở dữ liệu như sau:

- Dữ liệu có thể dễ dàng tích lũy lại.
- Dữ liệu có thể thao tác một cách thống nhất, tích hợp.
- Dữ liệu có thể được xử lý dễ dàng.

[126] (FAQ) Đã có những câu hỏi liên quan đến trung tâm dữ liệu trong các bài thi. Hãy cố gắng nắm chắc ý nghĩa của trung tâm dữ liệu, OLAP và OLTP.

[127] **Tác vụ chuẩn/không chuẩn** - **Standard/non-standard tasks**: Các tác vụ chuẩn là nhưng công việc mà thủ tục xử lý cố định ví dụ như các thủ tục kinh doanh hàng ngày và nhập dữ liệu bán hàng hàng ngày. Các tác vụ không chuẩn và những công việc mà thủ tục xử lý thay đổi theo từng trường hợp. Thí dụ tạo ra một tài liệu phân tích, cần các quá trình khác nhau tùy thuộc vào mục đích sử dụng, vì vậy nó được coi là tác vụ không chuẩn.

- Dữ liệu có thể được tìm kiếm dễ dàng.

## 2.5.3 Hệ thống đa phương tiện

| Chủ điểm | ➢ Các thí dụ về ứng dụng của đa phương tiện, bao gồm AI, 3DCG và nhận dạng mô hình.<br>➢ Các hệ thống ứng dụng đa phương tiên, bao gồm quảng bá Internet và VOD(Truyền hình theo yêu cầu). |
|---|---|

**Đa phương tiện - Multimedia** liên quan đến việc xử lý không chỉ ký tự và văn bản, mà còn là hỗn hợp của hình ảnh, video, âm thanh và các phương tiện truyền thông khác. Khái niệm này cũng ám chỉ tới các thiết bị và phần mềm sử dụng trong truyền thông đa phương tiện.

### Trí tuệ nhân tạo (AI)

Trí tuệ nhân tạo là hệ thống có thể thực hiện các quá trình suy diễn như hệ chuyên gia hoặc hệ thống dịch máy. Một máy tính chuẩn thường không làm gì hơn ngoài thao tác với số học và thực hiện các phép toán logic. AI, ngược lại, thực hiện các quá tình suy diễn trên các xâu ký tự. Hiện này Lisp và Prolog là những ngôn ngữ phù hợp nhất để phát triển phần mềm sử dụng công nghệ trí tuệ nhân tạo.

### Đồ họa 3 chiều (3D Computer Graphics, 3DCG)

Đồ họa 3 chiều trên máy tính[128] bao gồm tạo ra không gian ảo 3 chiều trong máy tính, đặt các mô hình vào không gian này, và di chuyển chúng. Công nghệ này cũng được sử dụng trong các bộ phim, trò chơi và hoạt hình. VRML[129] được sử dụng để phát triển 3DCG.

Bảng sau đây so sánh thực tại nhân tạo (artificial reality) và thực tại ảo (virtual reality).

| Loại | Giải thích |
|---|---|
| Thực tại nhân tạo (AR) | Công nghệ tạo ra một thế giới ảo bên trong máy tính, với cảm nhận của thực tại. Không cần công cụ đặc biệt nào để trải nghiệm thực tại nhân tạo này. |
| Thực tại ảo (VR) | Công nghệ tạo ra một thế giới tưởng tượng, có cảm nhận con người và cảm giác thế giới đó như là có thật.<br>Cần có màn hình 3 chiều riêng và thiết bị nhập liệu đặc biệt.<br>Thí dụ : phẫu thuật ảo từ xa, mô phỏng bay… |

[128] **Đồ họa máy tính** - **Computer graphics (CG)**: Công nghệ tạo ảnh sử dụng máy tính. Có những công cụ cho máy tính xử lý ảnh có sẵn hoặc là máy tính tự tạo ảnh mới (CGI – computer generated images).
[129] **Ngôn ngữ mô hình hóa thực tại ảo** - **VRML (virtual reality modeling language)**: ngôn ngữ đặc tả chuẩn 3DCG trên Internet.

# Các hệ thống ứng dụng đa phương tiện

Có nhiều hệ thống ứng dụng công nghệ đa phương tiện. Nhờ sự trợ giúp của WWW, chúng ta có thể dễ dàng truy cập đến một môi trường đồ họa, các nội dung đa phương tiện phong phú, sẵn sàng để sử dụng.

### Truyền hình Internet - Internet broadcasting

Công nghệ này ứng dụng đa phương trên nền Internet. Bằng việc sử dụng công nghệ truyền phát theo dòng[130](streaming distribution technology), các chương trình được phát đi trong thời gian thực. So với các phương pháp truyền thống, chi phí cho thiết bị ít hơn nhiều và hơn nữa có thể truyền đi toàn thế giới. Dịch vụ theo yêu cầu[131] cũng có thể triển khai trên nền công nghệ này vì vậy chúng ta có thể xem bất cứ khi nào, và chương trình gì chúng ta muốn.

Truyền hình Internet có nhiều định dạng khác nhau. Truyền hình theo yêu cầu lưu trữ nội dung trên một máy chủ phục vụ và phân phát chúng theo yêu cầu của người dùng . Truyền hình trực tiếp (Internet Live) phát các chương trình trực tiếp như các buổi hòa nhạc một cách đồng thời tới nhiều người dùng.

### Biên tập video phi tuyến - Non-linear editing

Đây là phương pháp biên tập video trong đó các hình ảnh được số hóa và sắp xếp vào đoạn video một cách tùy ý với sự trợ giúp của máy tính. Việc chỉnh sửa ảnh, thay đổi trật tự xuất hiện và tạo ra các phiên bản khác nhau khá dễ dàng. Theo phương pháp biên tập tuyến tính truyền thống, các hình ảnh, đoạn video phải được sắp xếp theo đúng thứ tự như khi đoạn phim đã hoàn tất.

### Truyền hình theo yêu cầu - Video on demand

Dịch vụ này gửi ngay một đoạn video qua một kênh truyền nào đó khi người xem có yêu cầu, thí dụ thông qua truyền hình cáp hai chiều(bidirectional CATV). Nhà cung cấp dịch vụ lưu trữ nhiều chương trình trên máy chủ phục vụ của họ và phân phát theo yêu cầu người xem.

Một máy chủ phục vụ truyền hình theo yêu cầu có thể đáp ứng được nhiều truy cập đồng thời của nhiều người xem, chương trình được yêu cầu sẽ

[130] **Truyền theo dòng** - **Streaming**: Công nghệ truyền dữ liệu cho phép nhận dữ liệu về và hiển thị ngay lập tức. Streaming cho phép truyền hình Internet phát nội dung đi và bên nhận hiển thị ngay mà không phải đợi thời gian trễ. Để streaming hoạt động tốt, băng thông của đường truyền phải cao hơn lưu lượng thông tin truyền đi/nhận về; tuy vậy, kết nối Internet thường chậm, do vậy dữ liệu thường được nén lại để đảm bảo thông tin được truyền phát theo thời gian thực. Việc xem một chương trình theo công nghệ truyền thốn thường mất thời gian vì dữ liệu phải tải về toàn bộ trước khi nó có thể được hiển thị. Với công nghệ streaming, việc hiển thị thông tin và tải dữ liệu sẽ được thực hiện cùng một lúc.

[131] **Theo yêu cầu** - **On-demand**: Là một tính năng của hệ thống cho phép cung cấp thông tin nào bất cứ khi nào được yêu cầu.

được phát lại từ đầu với từng yêu cầu truy cập đó. Do vậy, nhà cung cấp cần phải thiết lập một cơ sở dữ liệu hình ảnh và kết nối đến các thiết bị di động cuối, các thiết bị thu hình gia đình qua một kênh truyền băng thông rộng như dây cáp, không dây,...

**Câu hỏi**

**Q1** Intranet là gì?

**Q2** Trung tâm dữ liệu là gì?

**A1** Là mạng trong một công ty ứng dụng công nghệ của Internet. Thông thường một bức tường lửa được thiết lập giữa Internet và mạng của công ty để phòng chống việc rò rỉ thông tin của công ty ra bên ngoài.

**A2** Đó là một cơ sở dữ liệu của toàn bộ công ty, hỗ trợ việc ra quyết định. Khái niệm này b bao hàm việc lưu trữ một lượng lớn thông tin một cách có tổ chức để trợ giúp các quyết định kinh doanh. Đôi khi nó còn được gọi là cơ sở dữ liệu thông tin.

# Câu hỏi 1

**Q1.** Có một hệ thống quản lý tệp tin theo đơn vị là các khối. Mỗi khối có tám cung và mỗi cung lưu được 500 byte. Cần bao nhiêu cung để lưu trữ hai tệp tin, một tệp kích thước 2000 byte và tệp kia có kích thước 9000 byte? Bỏ qua các cung sử dụng bởi các cấu trúc thông tin quản lý, như thư mục.

a) 22 b) 26 c) 28 d) 32

# Đáp án câu 1

**Đáp án đúng: d**

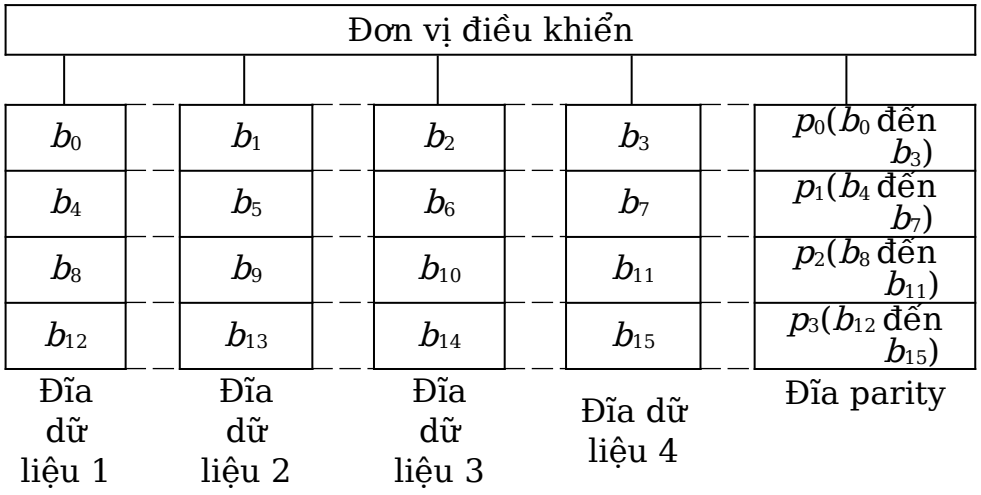
Tệp tin được lưu lại theo đơn vị kích thước là tám cung. Tám cung lưu được:
$8 * 500 = 4{,}000$ (byte)

Do vậy nếu một khối ít hơn 4000 byte, cả 4000 byte sẽ được sử dụng.
Tiếp theo chúng ta tìm số cung cần thiết để lưu trữ cho mỗi tệp tin 2000-byte và 9000-byte.

Dung lượng cần cho tệp tin 2000-byte là = 2,000 / 4,000= 0.5(khối)
→ cần 1 khối (= 8 cung)
Dung lượng cần cho tệp tin 9000-byte là = 9,000 / 4,000=2.25 (khối)
→ cần 3 khối (= 24 cung)
Vậy để lưu hai tệp tin trên , tổng số cung cần cấp phát là 8 + 24 = 32 cung.

## Câu hỏi 2

**Q2.** Thuật ngữ nào dưới đây chỉ quá trình chia nhỏ dữ liệu và lưu trữ trên nhiều đĩa cứng(xem hình dưới)? Ở đây, $b_0$ to $b_{15}$ tượng trưng cho thứ tự dữ liệu được lưu vào đĩa, theo đơn vị là bit, $p_0$ đến $p_3$ biểu diễn các bit parity sử dụng để phát hiện lỗi đĩa.




a) Striping
b) Disk caching
c) Blocking
d) Mirroring

## Đáp án câu 2

**Đáp án đúng: a**

**Stripping** là kỹ thuật phân phối một khối dữ liệu lên trên hai hoặc nhiều đĩa đồng thời. Nhờ stripping, mỗi khối dữ liệu có thể đọc và ghi song song trên nhiều đĩa, vì vậy tốc độ vào/ra tăng lên.
Stripping là một công nghệ của RAID. Như trên hình, cấu hình này có chứa một đĩa cứng riêng biệt cho việc kiểm tra parity; những cấu hình như vậy gọi là RAID2, RAID3 hoặc RAID4.

b) **Bộ đệm đĩa - Disk cache:** Là bộ nhớ đệm được đặt giữa đĩa cứng và bộ nhớ chính để cải thiện đáng kể tốc độ của đĩa cứng.

c) **Blocking** là cơ chế xử lý mỗi tập các bản ghi luân lý như một bản ghi vật lý (khối).

d) **Mirroring** là chuẩn bị nhiều đĩa cứng và ghi một cùng dữ liệu lên đồng thời các đĩa. Nếu một đĩa bị hỏng, thao tác vào ra vẫn được tiếp tục với những đĩa còn lại. Cấu hình này được gọi là RAID1.

# Câu hỏi 3

**Q3.** Thứ tự nào dưới đây là từ nhanh nhất đến chậm nhất theo nghĩa tốc độ truy cập bộ nhớ hiệu dụng.?

| | Bộ nhớ cache | | | Bộ nhớ chính |
|---|---|---|---|---|
| | Có cache hay không có cache? | Thời gian truy cập (ns) | Tỉ lệ trúng (%) | Thời gian truy cập (ns) |
| A | Không có cache | - | - | 15 |
| B | Không có cache | - | - | 30 |
| C | Có cache | 20 | 60 | 70 |
| D | Có cache | 10 | 90 | 80 |

a) A, B, C, D
b) A, D, B, C
c) C, D, A, B
d) D, C, A, B

# Đáp án câu 3

**Đáp án đúng: b**

Bộ nhớ cache ( bộ nhớ đệm) là bộ nhớ đặt giữa CPU và bộ nhớ chính để điều tiết sự chênh lệch tốc độ giữa hai thành phần trên. Tốc độ truy cập hiệu dụng có thể tăng lên bằng việc thêm vào một bộ nhớ đệm tốc độ cao và thực hiện đọc ghi lên bộ nhớ đệm nhiều nhất có thể.
Giả sử $t_c$ là tốc độ truy cập của bộ nhớ cache, $t_m$ là tốc độ truy cập của bộ nhớ chính, và h là tỉ lệ trúng. Tốc độ truy cập bộ nhớ hiệu dụng được tính như sau:

Tốc độ truy cập bộ nhớ hiệu dụng = $t_c * h + t_m * (1 - h)$

Tỉ lệ trúng là xác suất dữ liệu cần đọc đã có sẵn trong cache. Tỉ lệ trúng càng cao thì tốc độ truy cập bộ nhớ hiệu dụng càng nhanh.

Từ A đến D, chúng ta cần tính thời gian truy cập bộ nhớ hiệu dụng. Với trường hợp A và B, không có bộ nhớ cache, vì vậy thời gian truy cập bộ nhớ chính cũng là thời gian truy cập hiệu dụng. Ở đây chúng ta có thể coi h = 0. Dưới đây là kết quả theo thứ tự từ nhanh nhất đến chậm nhất ( thời gian truy cập hiệu dụng ngắn nhất đến dài nhất)

A: 15 (ns) (1)
B: 30 (ns) (3)
C: $0.6 * 20 + (1 - 0.6) * 70 = 40$ (ns) (4)
D: $0.9 * 10 + (1 - 0.9) * 80 = 17$ (ns) (2)

Vậy nếu ta sắp xếp từ A đến D theo thứ tự nhanh nhất đến chậm nhất, thứ tự sẽ là “A,D,B,C”.

## Câu hỏi 4

**Q4.** Hiện tượng phân mảnh file xảy ra khi sao chép một vài file từ thư mục này sang thư mục kia trên một đĩa cứng trong một máy tính cá nhân. Phát biểu nào dưới đây là đúng trong tình huống này.

a) Hiện tượng phân mảnh có để được loại bỏ bằng cách thực hiện kết xuất toàn bộ đĩa vật lý và sau đó phục hồi lại đĩa.
b) Thời gian truy cập thậm chí sẽ còn lâu hơn cho một vài file khác file bị phân mảnh.
c) Nếu file bị phân mảnh được sao chép lần nữa, tình trạng phân mảnh ở nơi sao chép đến sẽ còn tồi tệ hơn, nó sẽ không bao giờ bị loại bỏ.
d) Ngay cả khi phân mảnh xảy ra, kích thước của tệp cũng không bị thay đổi so với tệp gốc ban đầu.

## Đáp án câu 4

**Đáp án đúng: d**

Phân mảnh nghĩa là tệp không được ghi lên một vùng liên tục của đĩa, mà vào các vùng khác nhau trên đĩa. Khi có yêu cầu đến tệp bị phân mảnh, nhiều vùng khác nhau trên đĩa sẽ phải truy cập, do đó giảm hiệu suất xử lý. Tuy vậy kích thước tệp không bị thay đổi, chỉ đơn thuần bị chia nhỏ ra và lưu lại.

a) Nếu đĩa được sao chép một cách vật lý (nghĩa là từng bit thông tin trên đĩa giống hệt nhau cả về giá trị và vị trí trên đĩa), thì hiện tượng phân mảnh vẫn xảy ra. Nó phải được sao chép một cách luân lý.

b) Nếu một tệp được sao chép lên đĩa cứng mà đã có các tệp khác ở đó thì không tệp nào bị ảnh hưởng trừ tệp đang được sao chép. Các tệp khác không bị ảnh hưởng một cách vật lý.

c) Hiện tượng phân mảnh sẽ được giải quyết nếu tệp được sao chép đến nơi mới có kích thước vùng trống liên tiếp lớn hơn kích thước tệp.

# Câu hỏi 5

**Q5.** Bảng dưới đây chỉ ra thời gian xử lý cho một CPU và các thiết bị vào ra thực hiện 5 tác vụ độc lập. Tác vụ nào (a,b,c,d) có thể thực hiện đồng thời với "tác vụ có mức ưu tiên cao" để thời gian rỗi của CPU từ lúc bắt đầu thực hiện đến lúc kết thúc tác vụ đó là 0 ? Ở đây mỗi tác vụ sử dụng một thiết bị vào ra khác nhau và thực hiện song song. Tài nguyên dành cho hệ điều hành có thể bỏ qua.

Đơn vị: ms

| | Mức độ ưu tiên | Thời gian xử lý các tác vụ riêng rẽ |
|---|---|---|
| | Cao | CPU (3) →I/O (3) → CPU (3) → I/O (3) → CPU (2) |
| a) | Thấp | CPU (2) →I/O (5) → CPU (2) → I/O (2) → CPU (3) |
| b) | Thấp | CPU (3) →I/O (2) → CPU (2) → I/O (3) → CPU (2) |
| c) | Thấp | CPU (3) →I/O (2) → CPU (3) → I/O (1) → CPU (4) |
| d) | Thấp | CPU (3) →I/O (4) → CPU (2) → I/O (5) → CPU (2) |

# Đáp án câu 5

**Đáp án đúng:c**

Chúng ta kiểm tra hoạt động của mỗi tác vụ có mức ưu tiên thấp xem cái nào sử dụng CPU trong khi "tác vụ có mức ưu tiên cao" không sử dụng CPU. Các thiết bị vào ra là khác nhau với các tác vụ khác nhau, do vậy không phải đợi thao tác vào ra.

"Tác vụ có mức ưu tiên cao" sử dụng vào ra hai lần, mỗi lần 3ms. Nếu CPU được sử dụng bởi tác vụ có "mức ưu tiên thấp" chính xác 3ms, CPU sẽ không có thời gian rỗi.

Thí dụ, xem xét tác vụ (a). Trong thời gian 3ms khi "tác vụ có mức ưu tiên cao" sử dụng vào ra lần thứ nhất, tác vụ có mức ưu tiên thấp sẽ sử dụng CPU trong 2m. Trong 1 ms còn lại, cả 2 đều đợi cho thao tác vào ra hoàn tất, do vậy CPU sẽ không làm gì thời gian này.

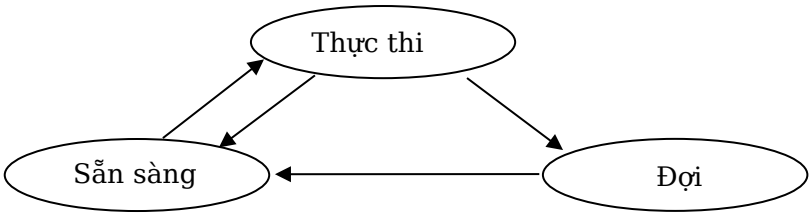
Tương tự như vậy, các tác vụ có "mức ưu tiên thấp" như (b) hay (d), có thời gian sử dụng 2ms, sẽ dẫn tới CPU nhàn rỗi. Ngược lại, tác vụ (c) sử dụng

4ms cuối cùng, ngay khi tác vụ có mức ưu tiên cao hoàn tất, do vậy sẽ không có thời gian nhàn rỗi cho CPU.

Vậy tác vụ có “mức ưu tiên thấp” có thể chạy cùng với “tác vụ có mức ưu tiên cao” để thời gian rỗi của CPU là bằng 0 là “CPU (3), I/O (2), CPU (3), I/O (1) , CPU (4) ”.

# Câu hỏi 6

**Q6.** Biểu đồ dưới đây mô tả quá trình chuyển trạng thái của một tác vụ (tiến trình) trên một hệ thống đa nhiệm. Khi nào thì tác vụ chuyển từ trạng thái thực thi sang trạng thái sẵn sàng ?

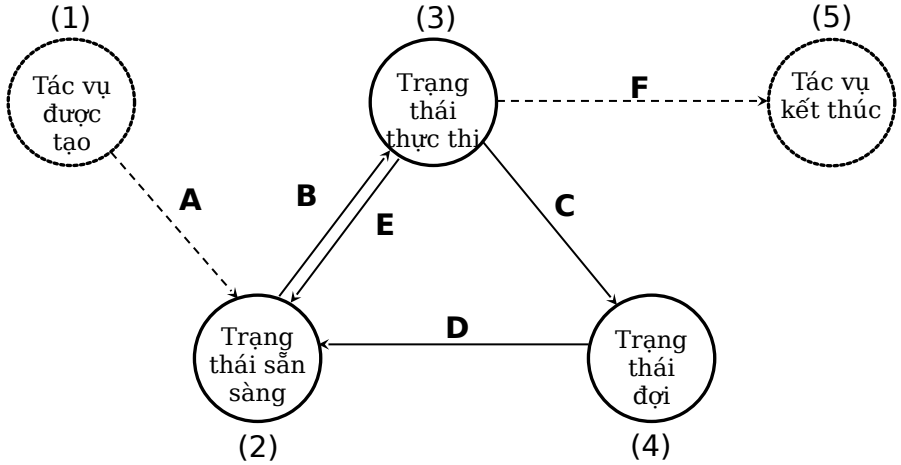



a) Một tác vụ có mức ưu tiên hơn đã chuyển sang trạng thái sẵn sàng.
b) Bộ lập lịch tạo ra một tác vụ mới.
c) Một thao tác vào ra hoàn tất.
d) Một thao tác vào ra vừa được yêu cầu.

# Đáp án câu 6

**Đáp án đúng: a**

Khi một tác vụ (tiến trình) được tạo, nó được chuyển sang trạng thái sẵn sàng và xếp vào hàng đợi. Sau đó, nó có chuyển sang trạng thái thực thi hay không tùy thuộc vào mức ưu tiên của tác vụ, hoặc chuyển sang trạng thái đợi khi có một thao tác vào ra xảy ra. Quá trình thay đổi như vậy được gọi là quá trình chuyển trạng thái của tác vụ.




Có thể giải thích các trạng thái và quá trình chuyển trạng thái như sau.

| STT | Trạng thái | Giải thích |
|---|---|---|

| | | |
|---|---|---|
| (1) | Tác vụ được tạo | Một tác vụ mới được đăng ký và xếp vào hàng đợi |
| (2) | Sẵn sàng | Tác vụ đợi được cấp phát tài nguyên CPU |
| (3) | Thực thi | Tác vụ được sử dụng tài nguyên CPU và đang thực hiện tính toán |
| (4) | Đợi | Tác vụ đợi một sự kiện hoàn tất, thí dụ, thao tác vào ra. |
| (5) | Tác vụ kết thúc | Tác vụ hoàn tất và nó không còn cần thiết nữa, sẽ được xóa khỏi hàng đợi. |

| Quá trình | Giải thích |
|---|---|
| A | Một tác vụ được tạo, nó được đưa vào hàng đợi |
| B | Tùy theo mức ưu tiên mà tác vụ được chuyển sang trạng thái thực thi |
| C | Do một sự kiện nào đó(đợi vào ra,...), tác vụ được chuyển sang trạng thái đợi |
| D | Chuyển sang trạng thái sẵn sàng sau khi một sự kiện hoàn tất. |
| E | Chuyển sang trạng thái sẵn sàng do tác vụ khác có mức ưu tiên cao hơn chuẩn bị thực thi. |
| F | Tác vụ đang thực thi hoàn tất. |

Như vậy việc chuyển trạng thái từ thực thi sang sẵn sàng là E. Điều này xảy ra khi một tác vụ có mức ưu tiên cao hơn chuyển sang trạng thái sẵn sàng.

b) Đây là quá trình chuyển trạng thái A trong hình.

c) Một thao tác vào ra xảy ra. Khi thao tác này hoàn tất, tác vụ được chuyển sang trạng thái sẵn sàng. Đây là quá trình chuyển trạng thái D trong hình.

d) Một thao tác vào ra vừa được tác vụ yêu cầu. Nếu điều này xảy ra, tác vụ sẽ gọi supervisor để thực hiện vào/ra sau đó nó được chuyển đến trạng thái đợi để chờ thao tác hoàn tất. Đây là quá trình chuyển trạng thái D trong hình.

## Câu hỏi 7

**Q7.** Phát biểu nào dưới đây phù hợp với môi hình hệ thống khách/chủ phục vụ ?

a) Máy khách và máy chủ phục phải chạy cùng loại hệ điều hành.
b) Máy chủ phục vụ gửi các yêu cầu xử lý dữ liệu, máy khách xử lý các yêu cầu đó.
c) Một máy chủ phục vụ có thể cho phép máy khách gửi yêu cầu xử lý đến máy chủ phục vụ khác nếu cần thiết.
d) Các chức năng phục vụ khác nhau phải đặt trên những máy tính khác nhau, thí dụ máy tính phục vụ tệp tin, máy tính phục vụ in ấn…

## Đáp án câu 7

**Đáp án đúng: c**

Mô hình hệ thống khách/chủ phục vụ gồm các thành phần gọi là khách, chủ phục vụ. Máy khách chỉ thực hiện vào ra dữ liệu, các quá trình khác thông qua máy chủ phục vụ, trong khi máy chủ phục vụ điều khiển tất các các thao tác vào/ra tùy theo phần cứng và kiểu máy chủ. Thông thường, Thành phần khách là các thiết bị có khả năng xử lý dữ liệu như máy tính cá nhân, trạm làm việc để có thể tự chạy ứng dụng. Thực tế, nó cũng thực hiện những công việc mà chỉ máy khách có thể làm được, như là hiển thị văn bản, vẽ hình. Ngược lại, máy chủ phục vụ truy cập cơ sở dữ liệu hay thực hiện in ấn theo các yêu cầu của máy khách. Hơn nữa, nếu một máy chủ phục vụ không thể đáp ứng được yêu cầu, nó có thể gửi yêu cầu đó đến máy chủ phục vụ khác. Lúc này, máy chủ phục vụ đầu tiên lại trở thành máy khách vì nó yêu cầu máy chủ khác thực hiện yêu cầu.

a) Các hệ điều hành khác nhau sẽ không gây ra vấn đề gì, miễn là chúng chạy cùng giao thức. Một máy chủ phục vụ chạy hệ điều hành UNIX có thể làm việc tốt với các máy khách chạy Windows.

b) Thứ tự phải là: máy khách gửi yêu cầu để xử lý, máy chủ xử lý yêu cầu đó.

d) Trong một hệ thông quy mô nhỏ, một máy chủ phục vụ và máy khách có thể chạy trên cùng một máy tính, thí dụ : X Windows của UNIX. Ngoài ra, sẽ không có hạn chế gì nếu máy khách và chủ phục vụ được xây dựng trên cùng một nền tảng (Hệ điều hành) và có thể nối với nhau qua mạng.

## Câu hỏi 8

**Q8.** Khi so sánh một hệ thống xử lý phân tán (bao gồm nhiều máy tính trên một diện rộng) với hệ thống xử lý tập trung, đâu là tính năng nổi bật nhất của hệ thống xử lý tập trung ?

a) Trong trường hợp thảm họa hoặc hỏng hóc, công việc phục hồi có thể tiến hành tập trung tại một nơi, tránh được nguy cơ phải dừng hệ thống trong một thời gian dài.
b) Khi hệ thống được quản lý tập trung, sẽ dễ dàng hơn trong việc thêm hoặc thay đổi các tính năng của hệ thống, hiếm khi công việc bị ùn tắc trì trệ.
c) Dữ liệu nhất quán và dễ dàng được quản lý, bảo trì thông qua việc cài đặt tập trung ở trung tâm.
d) Mặc dù vận hành và quản lý tài nguyên phần cứng, phần mềm trở nên phức tạp, nhưng việc mở rộng để tận dụng những ưu điểm của các nghệ mới khá dễ dàng.

## Đáp án câu 8

**Đáp án đúng: c**

Ý tưởng của một hệ thống xử lý tập trung là một máy tính đa năng cỡ lớn. Các thiết bị được tập trung tại một nơi nên việc quản lý trở nên dễ dàng. Tuy vậy nếu máy tính trung tâm bị hỏng, hệ thống sẽ có thể dừng hoạt động một thời gian dài.

a) Vì máy chủ thực hiện tất cả các xử lý, hệ thông sẽ bị ngừng hoạt động đến khi máy chủ được khôi phục. Nếu đây là lỗi trầm trọng, thời gian ngừng phục vụ của hệ thống sẽ có thể khá dài.
b) Trong hệ thống xử lý tập trung, máy chủ một mình phải xử lý tất các các yêu cầu. Nếu nội dung của các yêu cầu khác nhau đáng kể, máy chủ có thể không đáp ứng được tất cả những yêu cầu đó, dẫn tới hiện tượng đình trệ.

d) Một hệ xử lý tập trung xử lý tất cả các tác vụ tại máy chủ, vì vậy sẽ khá nặng nhọc cho máy chủ khi xử lý nhiều tác vụ doanh nghiệp riêng biệt. Hơn nữa, khi chúng ta muốn mở rộng hệ thống, sẽ có những tác vụ quan trọng không được phép dừng. Cùng với sự giới thiệu của công nghệ mới, một vài tác vụ có thể sẽ không được xử lý lâu hơn.

# Câu hỏi 9

**Q9.** Cho một công việc, công thức nào sau đây biểu diễn xấp xỉ mối quan hệ giữa thời gian quay vòng, thời gian CPU , thời gian Vào/Ra, và thời gian đợi của tiến trình ? Ở đây các loại thời gian phụ khác được bỏ qua.

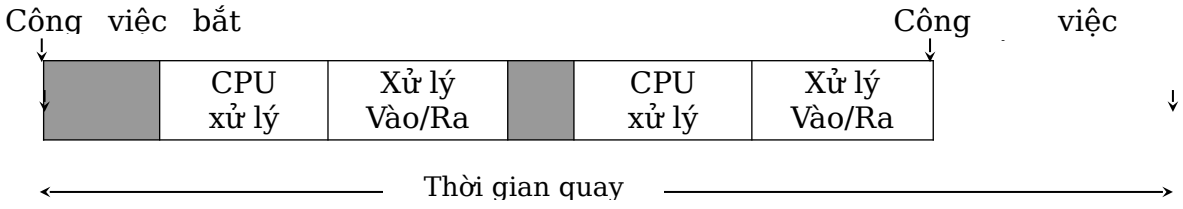
a) Thời gian đợi của tiến trình = Thời gian CPU + Thời gian quay vòng + Thời gian Vào/Ra.
b) Thời gian đợi của tiến trình = Thời gian CPU - Thời gian quay vòng + Thời gian Vào/Ra.
c) Thời gian đợi của tiến trình = Thời gian quay vòng - Thời gian CPU - Thời gian Vào/Ra.
d) Thời gian đợi của tiến trình = Thời gian Vào/Ra - Thời gian CPU - Thời gian quay vòng.

# Đáp án câu 9

**Đáp án đúng: c**

Thời gian đợi của tiến trình là thời gian tính đến đến khi CPU bắt đầu xử lý hoặc bắt đầu một thao tác Vào/Ra. Thời gian quay vòng(Turn-around time – TAT) là thời gian tính từ lúc bắt đầu gửi công việc cho đến khi nhận được kết quả. Khái niệm này chủ yếu dùng trong xử lý theo lô. TAT bao gồm cả thời gian CPU xử lý là thời gian Vào/Ra. Vậy thời gian đợi của tiến trình có thể tính bằng cách lấy TAT trừ đi thời gian CPU xử lý và thời gian Vào/Ra.

Trong hình dưới đây, vùng màu xáu diễn tả thời gian đợi của tiến trình.

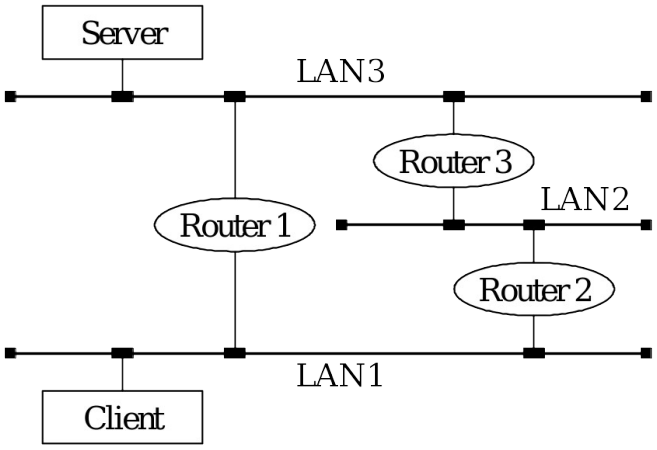



Như vậy, đối với một công việc, công thức biểu diễn mỗi liên hệ giữa giữa thời gian quay vòng, thời gian CPU , thời gian Vào/Ra, và thời gian đợi của tiến trình là :

Thời gian đợi của tiến trình = Thời gian quay vòng - Thời gian CPU - Thời gian Vào/Ra .

# Câu hỏi 10

**Q10.** Các mạng LAN được cài đặt như hình dưới đây. Máy chủ phục vụ kết nối tới LAN3, máy khách trên mạng LAN1 đang thực hiện các ứng dụng doanh nghiệp. Việc truyền dữ liệu giữa các mạng thường được thực hiện qua Router 1. Nếu có hư hỏng ở Router 1, Router 2 và 3 sẽ được sử dụng để truyền giữa mạng LAN1 và LAN3. Tỉ lệ sẵn sàng của thiết bị LAN kết nối LAN1 và LAN3 là bao nhiêu ? Ở đây, tỉ lệ hỏng hóc của mỗi bộ định tuyến là 0.1, không tính thời gian chuyển đổi giữa các thiết bị khi bị hư hỏng và giả định không có hỏng hóc nào trong LAN ngoại trừ các bộ định tuyến..

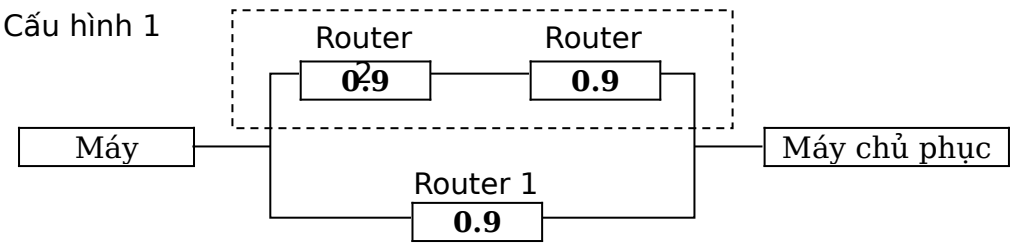



a) 0.729 b) 0.981 c) 0.990 d) 1.000

# Đáp án câu 10

**Đáp án đúng: b**

Để ý rằng “Router 1” và “kết nối nối tiếp giữa Router 2 và Router 3” mắc song song. Tỉ lệ hỏng của mỗi bộ định tuyến là 0.1 nên tỉ lệ sẵn sàng của chúng là 0.9. Do vậy, cấu hình mạng được vẽ lại như sau. Các giá trị trong hộp là tỉ lệ sẵn sàng của mỗi thiết bị.




Tỉ lệ sẵn sàng của cấu hình 1 = $0.9 * 0.9 = 0.81$

Tỉ lệ sẵn sàng của toàn hệ thống = $1 - (1 - 0.81) * (1 - 0.9) = 0.981$.

# Câu hỏi 11

**Q11.** Phát biểu nào dưới đây phù hợp với khái niệm thực tại ảo (VR – Virtual Reality)?

a) Sử dụng công nghệ như CG (đồ họa máy tính), thực tại ảo biểu diễn thế giới bên trong máy tính như thể đó là thế giới thật.
b) Với mục đích cái thiện giao diện người dùng, nó không hiển thị một bức ảnh dần dần từ trên xuống, mà đầu tiên hiển thị một bức ảnh mờ, kiểu như khảm, rồi rõ nét dần dần.
c) Thực tại ảo xác định các kết quả giả thuyết có thể đạt được hay không thông qua mô phỏng máy tính .Thí dụ: đường ống gió sử dụng trong các thiết kế ôtô và máy bay.
d) Các khả năng như nhân dang người, suy diễn có thể thực hiện được trên máy tính nhờ thực tại ảo.

# Đáp án câu 11

**Đáp án đúng: a**

Thực tại ảo (VR) dùng để tạo ra các cảm nhận thực tế một cách nhân tạo bằng việc kết hợp đồ họa máy tính và hiệu ứng âm thanh. Để giống với thực tế, người ta có thể sử dụng một màn hình đặc biệt như head-mounted (Head-mounted Display) cho việc hiển thị hình ảnh 3 chiều và thiết bị nhập đặc biệt như găng tay dữ liệu. Hơn nữa, các tác động tới hình ảnh cũng được phản hồi lại tới người xem, nâng cao tính thực tế của hình ảnh.

b) Phát biểu này mô tả kỹ thuật xen kẽ - interlace.

c) Phát biểu này mô tả kỹ thuật mô phỏng.

d) Phát biểu này mô tả công nghệ AI (Trí tuệ nhân tạo).